

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 21, 2004, 08:57:01 ; Search time 8.18182 Seconds

(without alignments)
528.928 Million cell updates/sec

Title: US-09-816-989A-2

Perfect score: 213

Sequence: 1 AKKYAKKAKAKAKAKAKA.....AKYKAAKAAKAAKAAAYEA 45

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : PIR 76:*

1: PIR1:*

2: PIR2:*

3: PIR3:*

4: PIR4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	103.5	48.6	210	A25550	histone H1 - sea u
2	98	46.0	394	P90725	membrane spanning
3	98	46.0	394	G85576	membrane spanning
4	98	46.0	421	JV0057	tolA protein - Bsc
5	94.5	44.4	239	AEI317	hypothetical prote
6	90	42.3	243	AEI689	hypothetical prote
7	89.5	42.0	1390	2	sperm tail-specifi
8	89.5	42.0	211	A28100	histone H1-beta, e
9	89.5	42.0	214	G70673	probable hupb - My
10	88	41.3	899	B87553	DNA topoisomerase
11	86.5	40.6	384	B43592	outer membrane pro
12	86	40.4	199	F70742	hypothetical prote
13	86	40.4	344	S34153	met101-1 protein -
14	86	40.4	347	E83525	TOLA protein PA097
15	85	39.9	182	S61926	histone H1 homolog
16	82	38.5	248	1	histone H1, gonada
17	82	38.5	262	2	histone H1, gonada
18	82	38.5	568	2	1-acylglycerol-3-p
19	81.5	38.3	311	2	asparaginyl-ctna s
20	80.5	37.8	277	2	hypothetical prote
21	80.5	37.8	445	2	probable NLP/P60 f
22	80.5	37.8	952	2	probable nucotin (i
23	80	37.6	194	1	probable DNA topoi
24	80	37.6	206	1	histone H1 - trout
25	80	37.6	291	1	histone H1 - rainb
26	79.5	37.3	217	2	hypothetical prote
27	79	37.1	62	2	histone H1-gamma,
28	79	37.1	205	2	hypothetical prote
29	79	37.1	244	2	cgr-1 protein - C
					histone H1 - midge

30	79	37.1	347	2	A81794	probable rotamase
31	78	36.6	217	2	S29309	hypothetical prote
32	78	36.6	231	2	S59589	histone H1 - Chlam
33	78	36.6	287	2	S45662	histone H1 - comat
34	78	36.6	309	2	G83013	polyhydroxyalkanoa
35	77.5	36.4	376	2	AG0592	tolA protein [impo
36	77	36.2	220	2	A28456	histone H1.10 - ch
37	77	36.2	225	1	S51660	histone H1-5 [vali
38	77	36.2	523	2	B38145	invariant surface
39	77	36.2	580	2	B70868	probable transfera
40	76	35.7	229	2	I51227	histone H1A - Altr
41	76	35.7	241	2	JN0748	histone H1-II - Vo
42	76	35.7	388	2	AC0138	TOLA colicin impor
43	75.5	35.4	220	2	JC5954	ribosomal protein
44	75.5	35.4	924	2	T06636	hypothetical prote
45	75	35.2	212	2	A28470	histone H1 - mouse

ALIGNMENTS

```
RESULT 1
A25550      histone H1 - sea urchin (Lytechinus pictus)
C:Species: Lytechinus pictus (painted urchin)
C>Date: 30-Jun-1988 #sequence_revision 30-Jun-1988 #text_change 23-Jul-1999
C:Accession: A25550
R:Knowles, J.A.; Childs, G.J.
Nucleic Acids Res. 14, 8121-8133, 1986
A>Title: Comparison of the late H1 histone genes of the sea urchins Lytechinus pictus an
A:Reference number: A25550; MUID:8704078; PMID:3022245
A:Accession: A25550
A:Molecule type: DNA
A:Residues: 1-210 <KNO>
A:Cross-references: GB:X04488; NID:99616; PIDN:CA28177.1; PID:99617
C:Superfamily: histone H1
C:Keywords: chromosomal protein; DNA binding; nucleosome; nucleu

Query Match      48.6%; Score 103.5; DB 2; Length 210;
Best Local Similarity 65.1%; Pred. No. 0.0053;
Matches 28; Conservative 3; Mismatches 11; Indels 1; Gaps 1;

Cy      1 AKKYAKKAKAEK-AKKAYKAAEKAKYKAAEKAAKAAKAA 42
Db      166 AKKAKKPAKKPAKKAKKPAKKAKKPAKKAKKPAKKAA 208

RESULT 2
P90725      membrane spanning protein Tola (imported) - Escherichia coli (strain O157:H7, substrain
C:Species: Escherichia coli
C>Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
Sasawara, N.; Yasunaga, T.; Kuwara, S.; Shiba, T.; Hatori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A>Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: P90725
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-394 <HAY>
A:Cross-references: GB:BA00007; PIDN:BA834197.1; PID:913360233; GSPDB:GN00154
A:Experimental source: Strain O157:H7, substrain RIMD 0509952
C:Genetics:
A:Gene: EC90774

Query Match      46.0%; Score 98; DB 2; Length 394;
Best Local Similarity 65.3%; Pred. No. 0.028;
Matches 32; Conservative 2; Mismatches 9; Indels 6; Gaps 3;

Cy      2 KKYAKKAAKAA---KK--AYKAAEKAAKAYEKAAEKAAKAAAYEA 45
Db      ||:||||| ||| ||| ||||| ||||| ||||| ||||| |||||
```


RESULT 7

S51364

A/Title: Drosophila-specific protein met101(2) - fruit fly (Drosophila hydei)

A/Reference number: A70500; MUID:98295987; PMID:9634230

A/Accession: S51364; #sequence_revision 26-Jul-1996 #text_change 17-Mar-1999

C/Date: 19-Jul-1996

A/Status: preliminary; nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-1390 <NEB>

A/Cross-references: EMBL:X73481

R/Neesen, J.; Heinlein, U.A.O.; Buemann, H.

Submitted to the EMBL Data Library, June 1993

A/Accession: S34153

A/Residues: 1-163, 'E', 164-236, 'Q', 237-254, 257-320, 'E', 321-1390 <NEW>

A/Cross-references: EMBL:X73481; NID:G313201; PID:G313202

C/Genetics:

A/Keywords: chromosome H1

A/Status: nucleic acid sequence not shown

A/Molecule type: DNA

A/Residues: 1-1390 <NEB>

A/Cross-references: EMBL:X73481

R/Neesen, J.; Heinlein, U.A.O.; Buemann, H.

Submitted to the EMBL Data Library, June 1993

A/Accession: S34153

A/Residues: 1-163, 'E', 164-236, 'Q', 237-254, 257-320, 'E', 321-1390 <NEW>

A/Cross-references: EMBL:X73481; NID:G313201; PID:G313202

C/Genetics:

A/Keywords: chromosome H1

A/Status: nucleic acid sequence not shown

A/Molecule type: DNA

A/Residues: 1-1390 <NEB>

A/Cross-references: EMBL:X73481

R/Neesen, J.; Heinlein, U.A.O.; Buemann, H.

Submitted to the EMBL Data Library, June 1993

A/Accession: S34153

A/Residues: 1-163, 'E', 164-236, 'Q', 237-254, 257-320, 'E', 321-1390 <NEW>

A/Cross-references: EMBL:X73481; NID:G313201; PID:G313202

C/Genetics:

A/Keywords: chromosome H1

A/Status: nucleic acid sequence not shown

A/Molecule type: DNA

A/Residues: 1-1390 <NEB>

A/Cross-references: EMBL:X73481

R/Neesen, J.; Heinlein, U.A.O.; Buemann, H.

Submitted to the EMBL Data Library, June 1993

A/Accession: S34153

A/Residues: 1-163, 'E', 164-236, 'Q', 237-254, 257-320, 'E', 321-1390 <NEW>

A/Cross-references: EMBL:X73481; NID:G313201; PID:G313202

C/Genetics:

A/Keywords: chromosome H1

A/Status: nucleic acid sequence not shown

A/Molecule type: DNA

A/Residues: 1-1390 <NEB>

A/Cross-references: EMBL:X73481

R/Neesen, J.; Heinlein, U.A.O.; Buemann, H.

Submitted to the EMBL Data Library, June 1993

A/Accession: S34153

A/Residues: 1-163, 'E', 164-236, 'Q', 237-254, 257-320, 'E', 321-1390 <NEW>

A/Cross-references: EMBL:X73481; NID:G313201; PID:G313202

C/Genetics:

A/Keywords: chromosome H1

A/Status: nucleic acid sequence not shown

A/Molecule type: DNA

A/Residues: 1-1390 <NEB>

A/Cross-references: EMBL:X73481

R/Neesen, J.; Heinlein, U.A.O.; Buemann, H.

Submitted to the EMBL Data Library, June 1993

A/Accession: S34153

A/Residues: 1-163, 'E', 164-236, 'Q', 237-254, 257-320, 'E', 321-1390 <NEW>

A/Cross-references: EMBL:X73481; NID:G313201; PID:G313202

C/Genetics:

A/Keywords: chromosome H1

A/Status: nucleic acid sequence not shown

A/Molecule type: DNA

us-09-816-989a-2.open.rpr

A/Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A/Reference number: A70500; MUID:98295987; PMID:9634230

A/Accession: S51364; #sequence_revision 26-Jul-1996 #text_change 17-Mar-1999

C/Date: 19-Jul-1996

A/Status: preliminary; nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-214 <COL>

A/Cross-references: GB:283018; GB:AL123456; NID:G3261671; PIDN:CA05427.1; PID:G1694845

A/Experimental source: strain H37Rv

C/Genetics:

A/Keywords: histone H1

A/Status: nucleic acid sequence not shown

A/Molecule type: DNA

A/Residues: 1-214 <COL>

A/Cross-references: GB:283018; GB:AL123456; NID:G3261671; PIDN:CA05427.1; PID:G1694845

A/Experimental source: strain H37Rv

C/Genetics:

A/Keywords: histone H1

A/Status: nucleic acid sequence not shown

A/Molecule type: DNA

A/Residues: 1-214 <COL>

A/Cross-references: GB:283018; GB:AL123456; NID:G3261671; PIDN:CA05427.1; PID:G1694845

A/Experimental source: strain H37Rv

C/Genetics:

A/Keywords: histone H1

A/Status: nucleic acid sequence not shown

A/Molecule type: DNA

A/Residues: 1-214 <COL>

A/Cross-references: GB:283018; GB:AL123456; NID:G3261671; PIDN:CA05427.1; PID:G1694845

A/Experimental source: strain H37Rv

C/Genetics:

A/Keywords: histone H1

A/Status: nucleic acid sequence not shown

A/Molecule type: DNA

A/Residues: 1-214 <COL>

A/Cross-references: GB:283018; GB:AL123456; NID:G3261671; PIDN:CA05427.1; PID:G1694845

A/Experimental source: strain H37Rv

C/Genetics:

A/Keywords: histone H1

A/Status: nucleic acid sequence not shown

A/Molecule type: DNA

A/Residues: 1-214 <COL>

A/Cross-references: GB:283018; GB:AL123456; NID:G3261671; PIDN:CA05427.1; PID:G1694845

A/Experimental source: strain H37Rv

C/Genetics:

A/Keywords: histone H1

A/Status: nucleic acid sequence not shown

A/Molecule type: DNA

A/Residues: 1-214 <COL>

A/Cross-references: GB:283018; GB:AL123456; NID:G3261671; PIDN:CA05427.1; PID:G1694845

A/Experimental source: strain H37Rv

C/Genetics:

A/Keywords: histone H1

A/Status: nucleic acid sequence not shown

A/Molecule type: DNA

A/Residues: 1-214 <COL>

A/Cross-references: GB:283018; GB:AL123456; NID:G3261671; PIDN:CA05427.1; PID:G1694845

A/Experimental source: strain H37Rv

C/Genetics:

A/Keywords: histone H1

A/Status: nucleic acid sequence not shown

A/Molecule type: DNA

A/Residues: 1-214 <COL>

A/Cross-references: GB:283018; GB:AL123456; NID:G3261671; PIDN:CA05427.1; PID:G1694845

A/Experimental source: strain H37Rv

C/Genetics:

A/Keywords: histone H1

A/Status: nucleic acid sequence not shown

A/Molecule type: DNA

A/Residues: 1-214 <COL>

A/Cross-references: GB:283018; GB:AL123456; NID:G3261671; PIDN:CA05427.1; PID:G1694845

A/Experimental source: strain H37Rv

C/Genetics:

A/Keywords: histone H1

RESULT 2
TOL A ECOLI STANDARD; PRT; 421 AA.
AC P19934;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE TOL A protein.
GN TOL A OR CIM OR EXCC OR LKY OR B0739.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxId=562;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=JMI05;
RX MEDLINE=90078104; PubMed=2687247;
RA Levensgood S.K., Webster R.E.;
RT "Nucleotide sequences of the tolA and tolB genes and localization of
RT their products, components of a multistep translocation system in
RT Escherichia coli."
RL J. Bacteriol. 171:6600-6609(1989).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
[3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97061202; PubMed=8905232;
RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kasahimoto K.,
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
RA Samei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horiiuchi T.;
RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 12.7-28.0 min region on the linkage map."
RL DNA Res. 3:137-155(1996).
[4]
RP DOMAINS.
RX MEDLINE=91296736; PubMed=2068069;
RA Levensgood S.K., Beyer W.F. Jr., Webster R.E.;
RT "TolA: a membrane protein involved in colicin uptake contains an
RT extended helical region."
RL Proc. Natl. Acad. Sci. U.S.A. 88:5939-5943(1991).
[5]
RP INTERACTION WITH PORINS.
RX MEDLINE=97133271; PubMed=8978668;
RA Derouiche R., Gavaioli M., Benedetti H., Prilipov A., Lazdunski C.,
RA Llobes R.;
RT "TolA central domain interacts with Escherichia coli porins."
RL EMBO J. 15:6408-6415(1996).
[6]
RP X-RAY CRYSTALLOGRAPHY (1.85 ANGSTROMS) OF 238-421.
RX MEDLINE=99332679; PubMed=10404600;
RA Lubkowski J., Hennecke F., Pluckethun A., Wlodawer A.;
RT "Fluorescent phage infection: crystal structure of g3p in complex
RT with its coreceptor, the C-terminal domain of TolA."
RL Structure 7:711-722(1999).
-1- FUNCTION: INVOLVED IN THE TONB-INDEPENDENT UPTAKE OF GROUP A
COLICINS (COLICINS A, E1, E2, E3, AND K). NECESSARY FOR THE
BINDING TO THE BACTERIA. ALSO INVOLVED IN THE TRANSLOCATION
OF BACTERIOPHAGE DNA.
-1- SUBUNIT: INTERACTS, VIA DOMAIN II, WITH PORINS OMPc, OMPc, PHOe
AND LAMB.

CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Inner membrane.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.ebi.ac.uk/announcements>
CC or send an email to license@ebi.ac.uk).
CC -----
DR EMBL: M28232; AAA24683.1; -
DR EMBL: AE000177; AAC73833.1; -
DR EMBL: D90713; BAA35405.1; -
DR PIR: J00057; J00057.
DR PDB: 1TOL; 20-MAY-99.
DR Ecocore: BGI1007; TOL A.
KW Transport; Protein transport; Bacteriocin transport; Transmembrane;
KW Repeat; Inner membrane; 3D-structure; Complete proteome.
FT DOMAIN 1 13 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 14 34 POTENTIAL. PERIPLASMIC (POTENTIAL).
FT DOMAIN 35 421 DOMAIN II (ALPHA-HELICAL).
FT DOMAIN 48 310 DOMAIN III (FUNCTIONAL).
FT DOMAIN 311 421 10 X TANDEN REPEATS OF [ED]-K(1,2)-
FT DOMAIN 224 278 A(2,4).
FT HELIX 335 349
FT TURN 350 351
FT TURN 353 354
FT HELIX 355 358
FT TURN 359 360
FT STRAND 363 369
FT TURN 371 372
FT STRAND 375 383
FT HELIX 385 397
FT HELIX 406 412
FT TURN 413 414
FT STRAND 416 421
SQ SEQUENCE 421 AA; 43156 MW; 8B2F52BA897C655E CRC64;
Query Match 46.0%; Score 98; DB 1; Length 421;
Best Local Similarity 65.3%; Pred. No. 0.0073;
Matches 32; Conservative 2; Mismatches 9; Indels 6; Gaps 3;
QY 2 KKYAKKAAEA---KK--AYKAAKAAAYEKAAKAAKAAEAAYAA 45
DB 247 KKAAKAAKAAEAADKKAARAAADKRAA-AKAAKAAKAAKAAEA 294
RESULT 3
ID MS72_DROHY STANDARD; PRT; 1391 AA.
AC Q08696;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Axoneme-associated protein msr101(2).
GN MS7101(2).
OS Drosophila hydei (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxId=7224;
[1]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RX MEDLINE=95045538; PubMed=7957199;
RA Neesen J., Padmanabhan S., Buenemann H.;
RT "Tandemly arranged repeats of a novel highly charged 16-amino-acid
RT motif representing the major component of the sperm-tail-specific
RT axoneme-associated protein family Dmsr101 form extended
RT alpha-helical rods within the extremely elongated spermatzoa of
RT Drosophila hydei."
RL Eur. J. Biochem. 225:1089-1095(1994).
-1- FUNCTION: POSSIBLE STRUCTURAL ROLE IN THE SPERM TAIL.

```

CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- TISSUE SPECIFICITY: TESTIS. PRIMARY SPERMATOCYTES AND EARLY
CC SPERMATIDS.
CC -1- DOMAIN: THE PREDOMINANT STRUCTURE IS ALPHA-HELICAL.
CC -1- POLYMORPHISM: LENGTH POLYMORPHISMS EXIST BETWEEN DIFFERENT
CC STRAINS, MOST LIKELY CAUSED BY LENGTH VARIATIONS WITHIN THE TANDEM
CC REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X73481; CAA51876.1; -
DR PIR; S51364; S51364.
DR FlyBase; FBgn0020733; Dhy1/msc101(2).
KM Sperm; Repeat; Multigene family; Polymorphism.
FT DOMAIN 332 1268 59 X 16 AA APPROXIMATE TANDEM REPEATS OF
SQ SEQUENCE 1391 AA; 159000 MW; 1B2A368F30F48878 CRC64;

Query Match 42.3%; Score 90; DB 1; Length 1391;
Best Local Similarity 57.8%; Pred. No. 0.13;
Matches 26; Conservative 4; Mismatches 11; Indels 4; Gaps 2;

OY 2 KKYAKAKAEK-ARKYAKAEAKK--AAKYAKAEAKAEAKA 42
ID DBH MYCSM STANDARD; PRT; 208 AA.
AC 09ZHC5;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE DNA-binding protein HU homolog (Histone-like protein) (Hlp).
GN HUP OR HLP.
OS Mycobacterium smegmatis.
OC Bacteria; Actinobacteriae; Actinomycetales;
OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1772;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=ATCC 700084 / mc(2)155;
RX MEDLINE=99110209; PubMed=9894918;
RX Lee B.H., Murugan-Oei B., Dick T.;
RA "Upregulation of a histone-like protein in dormant Mycobacterium
RA smegmatis.";
RL Mol. Genet. 260:475-479(1998).
CC -1- FUNCTION: THIS PROTEIN BELONGS TO THE HISTONE LIKE FAMILY OF
CC PROKARYOTIC DNA-BINDING PROTEINS WHICH ARE CAPABLE OF WRAPPING
CC DNA TO STABILIZE IT, AND PREVENT ITS DENATURATION UNDER EXTREME
CC ENVIRONMENTAL CONDITIONS (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE BACTERIAL HISTONE-LIKE PROTEIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF068138; AAD13809.1; -
DR HSSP; P02346; IHUJ.
DR InterPro; IPR000119; Bac_DNAbind.
DR Pfam; PF00216; Bac_DNA_binding; 1.
DR ProDom; PD000945; Bac_DNAbind; 1.

```

```

DR SMART; SM00411; BHL; 1.
DR PROSITE; PS00045; HISTONE_LIKE; 1.
KM DNA-binding; DNA condensation; Repeat.
FT DOMAIN 1 90 BACTERIAL HISTONE-LIKE DOMAIN.
FT DOMAIN 101 205 DEGENERATE REPEATS REGION.
SQ SEQUENCE 208 AA; 21230 MW; CA5F577F61F7EF09 CRC64;

Query Match 42.0%; Score 89.5; DB 1; Length 208;
Best Local Similarity 53.7%; Pred. No. 0.025;
Matches 29; Conservative 3; Mismatches 13; Indels 9; Gaps 2;

OY 1 AKKYAKAKAEK-----KKAYAAAKKAAKY--EKAAYEAKAEAKA 45
ID DB 111 AKKAAKAPAKKAAAKTATTAAKKAPAKKAAIKAPAKKATATAPAKKATKA 164

RESULT 5
HIB STRPU STANDARD; PRT; 211 AA.
AC P15669;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Histone H1-beta, late embryonic.
DE Strongylocentrotus purpuratus (Purple sea urchin).
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC Echinidea; Euechinidea; Echinoida; Strongylocentrotidae;
OC Strongylocentrotus.
OX NCBI_TaxID=7668;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88246461; PubMed=2837660;
RX Lai Z.-C., Childs G.;
RA "Characterization of the structure and transcriptional patterns of
RA the gene encoding the late histone subtype H1-beta of the sea urchin
RA Strongylocentrotus purpuratus.";
RL Mol. Cell. Biol. 8:1842-1844(1988).
CC -1- FUNCTION: HISTONES H1 ARE NECESSARY FOR THE CONDENSATION OF
CC NUCLEOSOME CHAINS INTO HIGHER ORDER STRUCTURES.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: BELONGS TO THE HISTONE H1/H5 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M20314; AAA30052.1; -
DR PIR; A28100; A28100.
DR HSSP; P02259; IHST.
DR InterPro; IPR005818; Histone_H1/H5.
DR InterPro; IPR003216; LinkerHist_N.
DR Pfam; PF00538; Linker_Histone; 1.
DR PRINTS; PR00624; HISTONEH5.
DR ProDom; PD000373; LinkerHist_N; 1.
DR SMART; SM00526; H15; 1.
KM Chromosomal protein, Nuclear protein; DNA-binding; Multigene family.
SQ SEQUENCE 211 AA; 22169 MW; 9F214581334ABE7A CRC64;

Query Match 42.0%; Score 89.5; DB 1; Length 211;
Best Local Similarity 61.9%; Pred. No. 0.026;
Matches 26; Conservative 3; Mismatches 10; Indels 3; Gaps 2;

OY 1 AKKYAKAKAEKAKKYAKAEAKKAAKAEKAEAKAEAKA 42
ID DB 139 SKKTKKVKKPKAPAKKAPKAAKAAK--KPAAPKAPAKKAA 177

RESULT 6

```

ID	DBH_MVCTU	STANDARD;	PRT;	214 AA.
AC	P95109;			
DT	15-JUL-1999	(Rel. 38, Created)		
DT	15-JUL-1999	(Rel. 38, Last sequence update)		
DT	16-OCT-2001	(Rel. 40, Last annotation update)		
DE	DNA-binding protein HU homolog (Histone-like protein) (Hlp) (21-kDa			
DE	lamini-2-binding protein).			
DN	HUP OR HLP OR LBP21 OR RV2986C OR MT3064 OR MTCY349.01.			
OS	Mycobacterium tuberculosis.			
OC	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;			
OC	Corynebacterineae; Mycobacteriaceae; Mycobacterium.			
OX	NCBI_TaxID=1773;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=H37Rv;			
RC	MEDLINE=968295987; PubMed=9634230;			
RA	Cole S.T., Broch R., Parkhill J., Garnier T., Churcher C., Harris D.,			
RA	Gordon S.V., Eiglmeier K., Gae S., Barry C.E. III, Tekala F.,			
RA	Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,			
RA	Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,			
RA	Holmesby T., Jergels K., Krogh A., McLean J., Moule S., Murphy L.,			
RA	Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,			
RA	Rutter S., Seeger K., Skelton S., Squares S., Squares R.,			
RA	Sulston J.E., Taylor K., Whitehead S., Barrett B.G.,			
RT	"Deciphering the biology of Mycobacterium tuberculosis from the			
RT	complete genome sequence.";			
RL	Nature 393:537-544 (1998).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CDC 1551 / Oshkosh;			
RA	Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,			
RA	Petersen J., Deboy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,			
RA	Kolony J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,			
RA	Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,			
RA	Bisler W.;			
RT	"Whole genome comparison of Mycobacterium tuberculosis clinical and			
RT	laboratory strains.";			
RL	Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.			
RN	[3]			
RP	SEQUENCE OF 71-86, AND DNA-BINDING.			
RC	STRAIN=H37Rv;			
RA	Prasad H.K., Annapurna P.S., Dey A.B., Tyagi J.S., Jain N.K.,			
RA	Saxena P.;			
RL	Submitted (DEC-1997) to the SWISS-PROT data bank.			
CC	-1- FUNCTION: THIS PROTEIN BELONGS TO THE HISTONE LIKE FAMILY OF			
CC	PROKARYOTIC DNA-BINDING PROTEINS WHICH ARE CAPABLE OF WRAPPING			
CC	DNA TO STABILIZE IT, AND PREVENT ITS DENATURATION UNDER EXTREME			
CC	ENVIRONMENTAL CONDITIONS (BY SIMILARITY).			
CC	-1- SIMILARITY: BELONGS TO THE BACTERIAL HISTONE-LIKE PROTEIN FAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation.			
CC	The European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
CC	EMBL; Z83018; CAB05427.1; -			
DR	EMBL; AE007412; AAK47393.1; -			
DR	PIR; G70673; G70673.			
DR	HSSP; P02346; 1HTU.			
DR	TIGR; MT3064; -			
DR	TubercuLib; RV2986C; -			
DR	InterPro; IPR000119; Bac_DNABind.			
DR	InterPro; IPR005819; Histone H5.			
DR	Pfam; PF00216; Bac DNA binding; 1.			
DR	PRINTS; PR00624; HISTONEH5.			
DR	Prodom; PD000945; Bac DNABind; 1.			
DR	SMART; SM00411; BHL_1.			
DR	PROSITE; PS00045; HISTONE_LIKE; 1.			
KW	DNA-binding; DNA condensation; Repeat; Complete proteome.			

```

FT DOMAIN 1 90 BACTERIAL HISTONE-LIKE DOMAIN.
TT SIGNAL 101 214 DEGENERATE REPEATS REGION.
SQ SEQUENCE 214 AA; 22187 MW; C609AP20RB353544 CRC64;

Query Match 42.0%; Score 89.5; DB 1; Length 214;
Best Local Similarity 55.6%; Pred. No. 0.026;
Matches 25; Conservative 4; Mismatches 13; Indels 3; Gaps 1;

OY 1 AKKAKKKAKKAKGATYKAAGAARAAKAAKAAKAAAYEA 45
Db 111 AKKAKKAPAKKATKAAKKAATKAPA--RKAATKAPAKKATATRA 152

RESULT 7
TMPB_TREPH STANDARD; PRT; 384 AA.
ID TMPB TREPH
AC P29720;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DB Treponema membrane protein B precursor (Antigen tmpb).
GN TMPB.
OS Treponema phagedenis.
OC Bacteria; Spirochaetes; Spirochaetales; Treponema.
OX NCBI_TaxId=162;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Kazan 5;
RX MEDLINE=91372983; PubMed=1894368;
RA Yelton D.B., Limberger R.J., Curci K., Malinovsky-Rummler F.,
RA Slivensky L., Schouls L.M., van Embden J.D., Charon N.W.;
RT "Treponema phagedensis encodes and expresses homologs of the Treponema
RL pallidum tmpa and tmpb proteins."
CC Infect. Immun. 59:3685-3693(1991).
CC -1 FUNCTION: TMP MAY SERVE AS A PORIN OR TRANSPORT PROTEIN FOR
CC LARGE MOLECULES.
CC -1 SUBCELLULAR LOCATION: OUTER MEMBRANE ASSOCIATED.
CC -1 SIMILARITY: TO TMPB OF T. PALLIDUM.
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC The European Bioinformatics Institute. There are no restrictions on its use by
CC non-profit institutions as long as its content is in no way modified and
CC modified and this statement is not removed. Usage by, and for commercial
CC entities requires a license agreement. See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; M58563; AAA27480.1; -.
DR PIR; B43592; B43592..
KW Antigen; Outer membrane; Repeat; Signal.
FT SIGNAL 1 21
FT CHAIN 22 384
FT FT DOMAIN 151 235
FT REPEAT 151 155
FT REPEAT 156 160
FT REPEAT 161 165
FT REPEAT 166 170
FT REPEAT 171 175
FT REPEAT 176 180
FT REPEAT 181 185
FT REPEAT 186 190
FT REPEAT 191 195
FT REPEAT 196 200
FT REPEAT 201 205
FT REPEAT 206 210
FT REPEAT 211 215
FT REPEAT 216 220
FT REPEAT 221 225
FT REPEAT 226 230
FT REPEAT 231 235
FT REPEAT 236 288
A-A-E.
6 X 6 AA TANDEM REPEATS OF [EA]-A-A-R-X-
```



```

CC -1- PTM: GLYCOSYLATED. GLYCOSYLATION MAY PROTECT THE PROTEIN FROM
CC PROTEOLYTIC DEGRADATION AND BE IMPORTANT FOR HEMAGGLUTINATION. IT
CC SUGGESTS THAT THE CARBOHYDRATE MOIETY MAY BE LOCATED WITHIN THE C-
CC TERMINAL DOMAIN OF HBHA.
CC -1- MISCELLANEOUS: SERUM FROM PATIENTS DIAGNOSED WITH ACTIVE
CC TUBERCULOSIS THAT HAD NOT BEEN VACCINATED CONTAINS ANTIBODIES THAT
CC RECOGNIZE HBHA, WHEREAS SERUM FROM HEALTHY INDIVIDUALS DOES NOT
CC CONTAIN ANY.
CC -1- SIMILARITY: STRONG, TO M. LEPRAE HBHA.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, A074390; AAC2652.1; -.
DR EMBL, 277162; CAB00936.1; -.
DR EMBL, A006951; AAK44716.1; ALT INIT.
DR PIR, F70742; F70742.
DR TIGR, MT0493; -.
DR Tuberculin; K00475; -.
DR Cell adhesion; Heparin-binding; Hemagglutinin; Glycoprotein;
DR Virulence; Complete proteome.
DR INIT MET 0
DR DOMAIN 151 193 ALA/LYS-RICH.
DR CONFLICT 120 120 R -> P (IN REF. 1).
DR SEQUENCE 198 AA; 21403 MW; 5137606F1EBB042 CRC64;
SQ
Query Match 40.4%; Score 86; DB 1; Length 198;
Best Local Similarity 65.6%; Pred. No. 0.053;
Matches 21; Conservative 4; Mismatches 7; Indels 0; Gaps 0;
QY 11 EKAKKAYRAAEKKAAYKAAKAAKAAKAAKAA 42
Db 160 KKAAPAKKAAPAKKAAPAKKAAPAKKAAPAKKA 191
RESULT 10
MS11_DROHY STANDARD; PRT; 344 AA.
AC Q08695;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Aconeme-associated protein msct101(1).
GN MS1101(1).
OS Drosophila hydei (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydriidae; Drosophilidae; Drosophila.
OC NCBI_Taxid=7224;
RN [1]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC TISSUE=Testis;
RX MEDLINE=94200512; PubMed=8150205;
RA Neesen J., Buemann H., Heinlein U.A.;
RA "The Drosophila hydei gene Dmsct101(1) encodes a testis-specific,
RA repetitive, aconeme-associated protein with differential abundance in
RA Y chromosomal deletion mutant flies."
RL Dev. Biol. 162:414-425 (1994).
RT
CC -1- FUNCTION: POSSIBLE STRUCTURAL ROLE IN THE SPERM TAIL. IT IS
CC ASSOCIATED WITH AXONEMAL STRUCTURES.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- TISSUE SPECIFICITY: TESTIS. LOCATED IN SPERMATOCYTES AND
CC SPERMATID BUNDLES.
CC -1- DOMAIN: THE PREDOMINANT STRUCTURE IS ALPHA-HELICAL.
CC -1- POLYMORPHISM: THE NUMBER OF REPEATS VARIES BETWEEN STRAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

```

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, X73480; CA51875.1; -.
DR PIR, S34153; S34153.
DR FlyBase; FBgn011816; Dhyd\msct101(1).
KW Sperm; Repeat; Multigene family.
FT DOMAIN 58 337
FT SEQUENCE 344 AA; 37793 MW; 24C65D2510387E2A CRC64;
SQ
Query Match 40.4%; Score 86; DB 1; Length 344;
Best Local Similarity 57.8%; Pred. No. 0.087;
Matches 26; Conservative 3; Mismatches 12; Indels 4; Gaps 2;
QY 2 KKVAKKAK-AEKAKKAYKAAEAKK---AAKYEKAAAEKAAKAAKAA 42
Db 69 KEAAEKKKCAEAAKKEKAAKKEKCAEAAKKEKAAEKKCAEAA 113
RESULT 11
TOLA_PSEAB STANDARD; PRT; 347 AA.
AC P50600;
DT 01-OCT-1996 (Rel. 34, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Tola protein.
GN TOLA OR PA0971.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OC NCBI_Taxid=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PAO;
RX MEDLINE=97113525; PubMed=8955385;
RA Dennis J.J., Lafontaine E.R., Sokol P.A.;
RA "Identification and characterization of the tolRA genes of
RA Pseudomonas aeruginosa."
RL J. Bacteriol. 178:7059-7068 (1996).
RN [2]
RP REVISIONS TO N-TERMINUS.
RA Duan K., Sokol P.A.;
RA Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PAO1;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle D.J., Lagrou M.,
RA Garber R.L., Goltry L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Paulsig I.T.,
RA Smitth K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sater M.H., Hancock R.E.W., Lory S., Olson M.V.;
RA "Complete genome sequence of Pseudomonas aeruginosa PAO1, an
RA opportunistic pathogen."
RL Nature 406:959-964 (2000).
RT
CC -1- FUNCTION: INVOLVED IN THE TONB-INDEPENDENT UPTAKE OF PROTEINS
CC (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Inner membrane
CC (Potential).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```



```

KW DNA-binding; DNA condensation; Repeat.
FT FT DOMAIN 1 90 BACTERIAL HISTONE-LIKE DOMAIN.
FT FT DOMAIN 101 205 DEGENERATE REPEATS REGION.
FT FT CONFLICT 199 199 A -> T (IN REF. 2).
SQ SEQUENCE 205 AA; 21262 MW; 19FC67885DFE648 CRC64;

Query Match 40.1%; Score 85.5; DB 1; Length 205;
Best Local Similarity 58.5%; Pred. No. 0.061;
Matches 24; Conservative 2; Mismatches 12; Indels 3; Gaps 1;

OY 1 AKKYAKKAKKAKKAYKAAEPAKKAYEKAALAKKAAKKA 41
DB 111 AKKVAKKAPAKKATKATKAKKATKAKA 148

RESULT 13
RS16_CORGL STANDARD; PRT; 165 AA.
AC OSNNX3;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 30S ribosomal protein S16.
GN RSP8 OR CGL2054.
OS Corynebacterium glutamicum (Brevibacterium flavum).
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Corynebacteriaceae; Corynebacterium.
OX NCBI_TaxID=1718;
RN [1]
RM SEQUENCE FROM N.A.
RC STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;
RA Nakagawa S.;
RL "Complete genomic sequence of Corynebacterium glutamicum ATCC 13032."
RL Submitted (MAY-2002) to the EMBL/Genbank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE S16P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb.ch).
CC -----
CC DR EMBL; AP005280; BAB99447.1; -.
CC DR HAMAP; MF_00385; -; 1.
CC DR InterPro; IPR000307; Ribosomal_S16.
CC DR Pfam; PF00886; Ribosomal_S16; 1.
CC DR ProDom; PD003791; Ribosomal_S16; 1.
CC DR TIGRFAMs; TIGR00002; S16; 1.
CC DR PROSITE; PS00732; RIBOSOMAL_S16; FALSE_NEG.
CC KW Ribosomal protein; Complete proteome.
CC SQ SEQUENCE 165 AA; 17837 MW; 61DD81961B8C0846 CRC64;

Query Match 38.5%; Score 82; DB 1; Length 165;
Best Local Similarity 62.9%; Pred. No. 0.11;
Matches 22; Conservative 4; Mismatches 7; Indels 2; Gaps 1;

OY 7 KAKAEKAKKAYKAAEPAKKAYEKAALAKKAAKKA 41
DB 120 EAITEKKKAREDEYKAEKA--EYKAAEKAAKAA 152

RESULT 14
HI_PARAN STANDARD; PRT; 248 AA.
AC P02256;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Histone H1, gonadal.
DE Parechinus angulosus (Angulate sea urchin).
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;

```

```

CC Echinoidea; Echinoidea; Echinacea; Echinoidea; Echinidae;
OC Parechinus.
OX NCBI TaxId=7658;
RN [1]
RP SEQUENCE OF 1-84:
RX MEDLINE=80156831; PubMed=6767609;
RA Strickland W.N., Strickland M., de Groot P.C., von Holt C.,
RA Wiltmann-Liebold B.;
RT "The primary structure of histone H1 from sperm of the sea urchin
RT Parechinus angulosus. 1. Chemical and enzymatic fragmentation of the
RT protein and the sequence of amino acids in the four N-terminal
RT cyanogen bromide peptides."
RL Eur. J. Biochem. 104:559-566 (1980).
RN [2]
RX SEQUENCE OF 80-248.
RX MEDLINE=80156832; PubMed=7363905;
RA Strickland W.N., Strickland M., Brandt W.F., von Holt C., Lehmann A.,
RA Wiltmann-Liebold B.;
RT "The primary structure of histone H1 from sperm of the sea urchin
RT Parechinus angulosus. 2. Sequence of the C-terminal CNBr peptide and
RT the entire primary structure."
RL Eur. J. Biochem. 104:567-578 (1980).
CC -1- FUNCTION: HISTONES H1 ARE NECESSARY FOR THE CONDENSATION OF
CC NUCLEOSOME CHAINS INTO HIGHER ORDER STRUCTURES.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- TISSUE SPECIFICITY: Sperm.
CC -1- SIMILARITY: BELONGS TO THE HISTONE H1/H5 FAMILY.
DR HSSP; P02259; 1HST.
DR InterPro; IPR005818; Histone_H1/H5.
DR InterPro; IPR005819; Histone_H5.
DR InterPro; IPR003216; Linkerhist_N.
DR Pfam; PF00538; Linker histone; 1.
DR PRINTS; PR00624; HISTONEH5.
DR ProDom; PD000373; Linkerhist_N; 1.
DR SMART; SM00526; H15; 1.
DR Chromosomal protein; Nuclear protein; DNA-binding; Multigene family;
DR Sperm.
KW VARIANT 144 144 K -> R.
SQ SEQUENCE 248 AA; 26387 MW; 1B25B3F136541947 CRC64;

Query Match 38.5%; Score 82; DB 1; Length 248;
Best Local Similarity 57.8%; Pred. No. 0.16;
Matches 26; Conservative 3; Mismatches 14; Indels 2; Gaps 2;

OY 2 KKYAKKAKAKAKAKYKAAEAKKAKYKAAEAKKAA-AKEAAYEA 45
DB 120 KKAATTSAAAKAKA-KAAAKKAKAKAKAAKKAALAKKAAA 163

RESULT 15
HIE CHIPA STANDARD: PRT; 235 AA.
AC P40762;
DT 01-FEB-1995 (rel. 31, Created)
DT 01-FEB-1995 (rel. 31, Last sequence update)
DT 15-JUL-1999 (rel. 38, Last annotation update)
DE Histone H1B.
OS Chironomus pallidivittatus (Midge).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Nematocera; Chironomidae;
OC Chironomidae; Chironominae; Chironomus.
OX NCBI TaxId=7151;
RN [1]
RP SEQUENCE FROM N.A.
RA Schilze E., Wasniewski J.R., Nagel S., Gavenis K., Grossbach U.;
RL Submitted (XXL-1994) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: HISTONES H1 ARE NECESSARY FOR THE CONDENSATION OF
CC NUCLEOSOME CHAINS INTO HIGHER ORDER STRUCTURES.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: BELONGS TO THE HISTONE H1/H5 FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib.ch).
CC -----
DR EMBL; U29106; AAA21714.1; -.
DR HSSP; P08287; 1GHC.
DR InterPro; IPR005818; Histone_H1/H5.
DR InterPro; IPR005819; Histone_H5.
DR InterPro; IPR003216; Linkerhist_N.
DR Pfam; PF00538; Linker histone; 1.
DR PRINTS; PR00624; HISTONEH5.
DR ProDom; PD000373; Linkerhist_N; 1.
DR SMART; SM00526; H15; 1.
DR Chromosomal protein; Nuclear protein; DNA-binding; Multigene family.
SQ SEQUENCE 235 AA; 24446 MW; 340BC5B9A85002AC CRC64;

Query Match 38.0%; Score 81; DB 1; Length 235;
Best Local Similarity 41.3%; Pred. No. 0.19;
Matches 26; Conservative 2; Mismatches 17; Indels 18; Gaps 1;

OY 1 AKKYAKKAKAKA-----KKAAYKAAEAKKAKYKAAEAKKAAKAA 42
DB 126 AKQVDKPKKAPAPKPKSTNKRVTGKKVKKPKPAKKPEAKKATYAKAPAKKAAKPA 185

OY 43 YEA 45
DB 186 KKA 188

```

Search completed: January 21, 2004, 09:00:58
 Job time : 8.13636 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 21, 2004, 08:53:31 ; Search time 18.1169 Seconds
(without alignments)
640,969 Million cell updates/sec

Title: US-09-816-989A-2

Perfect score: 213
Sequence: 1 AKYKAKKAKKAKKAKKAA.....AKYKAAKRAAKAAAYEA 45

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 23:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_nhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteriopl:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	98	46.0	394	16	Q8X965
2	98	46.0	421	16	Q8FJ71
3	94.5	44.4	239	16	Q8V5M4
4	91	42.7	244	16	Q8AJX2
5	90	42.3	243	16	Q8ZAE7
6	88.5	41.5	275	5	Q01395
7	88.5	41.5	372	2	Q9WXX1
8	88	40.3	899	16	Q9ASJ6
9	86.5	40.6	200	16	Q8XVW7
10	86	40.4	155	16	Q8PI40
11	86	40.4	168	16	Q8F930
12	86	40.4	212	3	Q8J946
13	85	39.9	98	5	Q8WQ44
14	85	39.9	111	5	Q8TR93
15	85	39.9	182	2	Q45370
16	84	39.4	60	5	Q9U3W3

17	84	39.4	144	13	Q8AW13	Q8AW13 oncorhynch
18	83.5	39.2	483	12	Q8OND1	Q8OND1 ectocarpus
19	83	39.0	293	10	Q9AT18	Q9AT18 lens culina
20	83	39.0	389	16	Q9CM70	Q9CM70 pasteurella
21	82.5	38.7	243	5	Q23784	Q23784 chironomus
22	82	38.5	190	5	Q8MYC3	Q8MYC3 mytilus cal
23	82	38.5	262	16	Q9S2M2	Q9S2M2 streptomyce
24	82	38.5	568	3	Q94567	Q94567 schizosacch
25	81.5	38.3	156	16	Q8PEU0	Q8PEU0 xanthomonas
26	81.5	38.3	311	12	Q84528	Q84528 parametium
27	81.5	38.3	482	2	Q93LK4	Q93LK4 enterococcu
28	81	38.0	81	5	Q9NFP6	Q9NFP6 trypanosoma
29	81	38.0	112	5	Q9XYX5	Q9XYX5 leishmania
30	81	38.0	241	5	Q23790	Q23790 chironomus
31	81	38.0	290	10	Q9AT24	Q9AT24 pium sattiv
32	81	38.0	295	10	Q8LKH9	Q8LKH9 pium fulvu
33	81	38.0	295	10	Q8LKH0	Q8LKH0 pium sattiv
34	81	38.0	295	10	Q9AT22	Q9AT22 lathyrus sa
35	81	38.0	295	10	Q9ZK20	Q9ZK20 pium sattiv
36	81	38.0	295	10	Q9AT25	Q9AT25 pium sattiv
37	81	38.0	297	10	Q9SX08	Q9SX08 pium sattiv
38	81	38.0	298	10	Q8LKH1	Q8LKH1 lathyrus ap
39	81	38.0	301	10	Q9AT23	Q9AT23 pium sattiv
40	81	38.0	306	10	Q9AT21	Q9AT21 lathyrus sa
41	80.5	37.8	277	16	Q9XAO3	Q9XAO3 streptomyce
42	80.5	37.8	445	3	Q9P308	Q9P308 neurospora
43	80	37.6	66	5	Q95QZ0	Q95QZ0 caenorhabdi
44	80	37.6	71	5	Q9NF08	Q9NF08 trypanosoma
45	80	37.6	81	5	Q9N6L9	Q9N6L9 trypanosoma

ALIGNMENTS

RESULT 1

Q8X965 PRELIMINARY; PRT; 394 AA.

AC Q8X965

DT 01-MAR-2002 (TREMBLrel. 20, Created)

DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)

DE Membrane spanning protein, required for outer membrane integrity

DE (Membrane spanning protein Tola).

GN TOLA OR Z0907 OR EGS0774.

OS Escherichia coli O157:H7.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Escherichia.

OX NCB1_TaxID=83334;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=O157:H7 / EDL933 / ATCC 700927;

RX MEDLINE=21074935; PubMed=11206551;

RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glaesner J.D., Rose D.J., Mayhew G.P., Evans P.S., Gregor J., Kirkpatrick H.A., Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L., Grobbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamianos K., Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C., Welch R.A., Blattner F.R.;

RA "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.;"

RL Nature 409:529-533(2001).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=O157:H7 / RIMD 0509952;

RX MEDLINE=21156231; PubMed=11258796;

RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K., Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T., Iida T., Takami H., Hattori M., Shinagawa H., Ogasawara N., Yasunaga T., Kuhara S., Shibata T., Hattori M., Shinagawa H.,

RT "Complete genome sequence of enterohaemorrhagic Escherichia coli O157:H7 and genomic comparison with a laboratory strain K-12.;"

RL DNA Res. 8:11-22(2001).

DR EMBL; AF005252; AAC55075.1; -

DR EMBL; AF005253; BAB34197.1; -

```

DR InterPro; IPR000104; Antifreeze_1.
DR PRINTS; PR00308; ANTIFREEZE1.
DR Complete proteome.
SQ SEQUENCE 394 AA; 40517 MW; 5B58D9E8230BDE28 CRC64;

Query Match
Best Local Similarity 46.0%; Score 98; DB 16; Length 394;
Matches 32; Conservative 2; Mismatches 9; Indels 6; Gaps 3;

OY 2 KKYAKKAAEKA---KK--AYKAAEKKAAYEKAAYEKAAYE 45
DB 220 KKAEEKAAEKAADKKAEEKAAEKAADKKAEEKAAEKAAYE 267

RESULT 2
OBFUTL1
AC OBFUTL1; PRELIMINARY; PRT; 421 AA.
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DE TOLA protein.
GN TOLA OR C0818.
OS Escherichia coli O6.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxId=21792;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=06:HI / CFT073 / ATCC 700928;
RX MEDLINE=2238234; PubMed=12471157;
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA Rasco D., Buckles E.L., Lion S.-R., Boutin A., Hackett J., Stroud D.,
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT "Extensive mosaic structure revealed by the complete genome sequence
RT of uropathogenic Escherichia coli."
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
DR EMBL; AB016757; F016757.1;
KM Complete proteome.
SQ SEQUENCE 421 AA; 43184 MW; DB296626F056D385 CRC64;

Query Match
Best Local Similarity 46.0%; Score 98; DB 16; Length 421;
Matches 32; Conservative 2; Mismatches 9; Indels 6; Gaps 3;

OY 2 KKYAKKAAEKA---KK--AYKAAEKKAAYEKAAYEKAAYE 45
DB 247 KKAEEKAAEKAADKKAEEKAAEKAADKKAEEKAAEKAAYE 294

RESULT 3
OBFUTL1
AC OBFUTL1; PRELIMINARY; PRT; 239 AA.
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DE Hypothetical protein lmo1941.
GN LMO1941.
OS Listeria monocytogenes.
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OX NCBI_TaxId=1639;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EGD-e / Serovar 1/2a;
RX MEDLINE=21537279; PubMed=11679669;
RA Glaeser P., Fiangeul L., Buchrieser C., Rusniok C., Amend A.,
RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
RA Charbit A., Chetoui A., Couve E., de Daruvar A., Delhoux P.,
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dusserget O.,
RA Entlan K.-D., Fehli H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Haut J., Jackson D.,

```

```

RA Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kurupak G.,
RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA Nordstiek G., Novella S., de Pablo B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schlueter T., Simoes N., Tixeret A.,
RA Vazquez-Boland J.-A., Voss H., Wehland J., Coesart P.,
RT "Comparative genomics of Listeria species."
RL Science 294:849-852(2001).
DR EMBL; AL591981; CAD00019.1; -
DR Listeria; LMO1941; -
DR InterPro; IPR002482; LysM.
DR Pfam; PF01476; LysM; 1.
DR SMART; SM00257; LysM; 1.
KM Hypothetical protein; Complete proteome.
SQ SEQUENCE 239 AA; 25836 MW; 72B59D576E0D7832 CRC64;

Query Match
Best Local Similarity 44.4%; Score 94.5; DB 16; Length 239;
Matches 27; Conservative 7; Mismatches 9; Indels 5; Gaps 2;

OY 2 KKYAKKAAEKA---AYKAAEKKAAYEKAAYEKAAYE 44
DB 124 KKAEEKAAEKAADKKAEEKAAEKAADKKAEEKAAEKAAYE 171

RESULT 4
OBFUTL2
AC OBFUTL2; PRELIMINARY; PRT; 244 AA.
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DE Putative secreted protein.
GN SCO1805 OR SCI33.04.
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteria; Actinobacteriales; Actinomycetales;
OC Streptomycetaceae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxId=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RX MEDLINE=21996410; PubMed=12000953;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kleser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kleser T., Larke L., Larke L., Murphy K., O'Neill S.,
RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward V., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2)."
RL Nature 417:141-147(2002).
DR EMBL; AL939110; CAC28545.1; -
KM Complete proteome.
SQ SEQUENCE 244 AA; 25524 MW; 61999D62CA23A7B0 CRC64;

Query Match
Best Local Similarity 42.7%; Score 91; DB 16; Length 244;
Matches 25; Conservative 4; Mismatches 11; Indels 2; Gaps 1;

OY 5 AKKAAEKA---AYKAAEKKAAYEKAAYEKAAYE 44
DB 78 SQVAAIEKKAAYEKAAYEKAAYEKAAYEKAAYE 119

RESULT 5
OBFUTL1
AC OBFUTL1; PRELIMINARY; PRT; 243 AA.
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)

```

```

DE Hypothetical protein lin2055.
GN LIN2055.
OS Listeria innocua.
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OX NCBI_TaxID=1642;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CLIP 11262 / Serovar 6a;
RX PubMed=11679669;
RA Glaeser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
BA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
RA Charbit A., Checrouni F., Couve E., de Daruvar A., Deloux P.,
Doman E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA Entian K.-D., Feibi H., Garcia-del Portillo F., Garrido P.,
Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
Jones L.-M., Kaerst U., Kretz U., Kunz M., Kunst F., Kurapat G.,
RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA Nordstedt G., Novella S., de Pablo B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schluter T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RT "Comparative genomics of Listeria species.";
RL Science 294:849-852(2001).
DR EMBL; AL596170; CAC97285.1; -.
DR Listlist; LIN02055; -.
DR InterPro; IPR002482; LYSM.
DR Pfam; PF01476; LYSM; 1.
DR SMART; SM00257; LysM; 1.
KM Hypothetical protein; Complete proteome.
SQ SEQUENCE 243 AA; 25963 MW; 6B2493D143B159D1 CRC64;

Query Match
Best Local Similarity 42.3%; Score 90; DB 16; Length 243;
Matches 30; Conservative 3; Mismatches 11; Indels 22; Gaps 2;

Qy 1 AKKTA-KKAKAKKAYKAAEKKA-----AKYKAAAEKAAA 38
Db 106 AKKAAEEKAAAEKAAEKKAADKSGDEAKAAAKKQEAEEKAAAEKAAA 165
Qy 39 KEAYE 44
Db 166 DKAKE 171

RESULT 6
ID 001395 PRELIMINARY; PRT; 275 AA.
AC 001395;
DT 01-JUL-1997 (TREMblrel. 04, Created)
DT 01-JUL-1997 (TREMblrel. 04, Last sequence update)
DT 01-OCT-2002 (TREMblrel. 22, Last annotation update)
DE Akonene-associated protein MST101(3).
GN MST101(3) OR DHMST101.
OS Drosophila hydei (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7224;
RN [1]
RP SEQUENCE FROM N.A.
RA Neesen J., Heinlein U.A.O., Buenemann H.;
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: POSSIBLE STRUCTURAL ROLE IN THE SPERM TAIL (BY
CC SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC -1- TISSUE SPECIFICITY: TESTIS (BY SIMILARITY).
CC -1- DOMAIN: THE PREDOMINANT STRUCTURE IS ALPHA-HELICAL.
DR EMBL; U85627; AAB51369.1; -.
DR FlyBase; FBgn0020733; dhyalmer101(3).
KM Sperm; Repeat; Multigene family.
FT DOMAIN 64 255
FT 13 X 16 AA APPROXIMATE TANDEM REPEATS OF
FT X-[KQ]-K-C-[AE]-E-X-A-[X]-K-X-X-X-X-
FT [AE]-X.
SQ SEQUENCE 275 AA; 30436 MW; 76BAA7B2ADDF32C CRC64;

```

```

Query Match
Best Local Similarity 41.5%; Score 88.5; DB 5; Length 275;
Matches 25; Conservative 1; Mismatches 12; Indels 3; Gaps 1;

Qy 5 AKKAAEKAKKAYKAAEKKA-----AAKYKAAAEKAAAEKAA 42
Db 63 AKKCAAEKAEKAEKAAEKKAADKSGDEAKAAAKKQEAEEKAAAEKAAA 103

RESULT 7
ID 09WMX1 PRELIMINARY; PRT; 372 AA.
AC 09WMX1;
DT 01-NOV-1999 (TREMblrel. 12, Created)
DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)
DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
DE TOLA protein.
GN TOLA.
OS Pseudomonas putida.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=303;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=mt-2;
RX MEDLINE=96198174; PubMed=8626299;
RA Rodriguez-Herva J.U., Ramos-Gonzalez M.I., Ramos J.;
RT "The Pseudomonas putida peptidoglycan-associated outer membrane
lipoprotein (PAL) is involved in maintenance of the integrity of the
cell envelope.";
RL J. Bacteriol. 178:1699-1706(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=mt-2;
RA Ramos-Gonzalez I.;
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=mt-2;
RA Rodriguez-Herva J.U.;
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=mt-2;
RX MEDLINE=96422022; PubMed=8824639;
RA Rodriguez-Herva J.U., Ramos J.;
RT "Characterization of an OprL null mutant of Pseudomonas putida.";
RL J. Bacteriol. 178:5836-5840(1996).
DR EMBL; X74218; CAB50780.1; -.
DR InterPro; IPR005819; Histone_H5.
DR InterPro; IPR006260; TonB_C.
DR PRINTS; PRO0624; HISTONEH5.
DR TIGRPMs; TIGR01352; tonB_Cterm; 1.
SQ SEQUENCE 372 AA; 40133 MW; 87F49785EC3C0BC CRC64;

Query Match
Best Local Similarity 41.5%; Score 88.5; DB 2; Length 372;
Matches 27; Conservative 8; Mismatches 10; Indels 23; Gaps 1;

Qy 1 AKKTA-KKAKAKKAYKAAEKKA-----AKKAA 37
Db 121 AEDAKAAAEKAAAEKKAADKSGDEAKAAAKKQEAEEKAAAEKAAA 180
Qy 38 AKAAAYE 45
Db 181 AEAEEKKA 188

RESULT 8
ID 09A5J6 PRELIMINARY; PRT; 899 AA.
AC 09A5J6;

```

```

DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE DNA topoisomerase (EC 5.99.1.2).
GN CC2451.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacteriales;
OC Caulobacteraceae; Caulobacter.
OX NCBI_TaxID=155892;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19089 / CB15;
RX MEDLINE=21173698; PubMed=11259647;
RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA Kolonay J.F., Smit J., Craen M.B., Khouli H., Shetty J., Berry K.,
RA Ueberback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RA "Complete genome sequence of Caulobacter crescentus.";
RA Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
CC -1- FUNCTION: THE REACTION CATALYZED BY TOPOISOMERASE LEADS TO THE
CC -1- CONVERSION OF ONE TOPOLOGICAL ISOMER OF DNA TO ANOTHER (BY
CC SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP-INDEPENDENT BREAKAGE OF SINGLE-STRANDED
CC DNA, FOLLOWED BY PASSAGE AND REJOINING.
CC -1- MISCELLANEOUS: WHEN A TOPOISOMERASE TRANSIENTLY BREAKS A DNA
CC BACKBONE BOND, IT SIMULTANEOUSLY FORMS A PROTEIN-DNA LINK, IN
CC WHICH A TYROSYL OXYGEN IN THE ENZYME IS JOINED TO A DNA PHOSPHORUS
CC AT ONE END OF THE ENZYME-SEVERED DNA STRAND (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO PROKARYOTIC TYPE I/III TOPOISOMERASE
CC FAMILY.
CC DR EMBL: AE005914; AK24422.1; -.
CC DR HSS: P06612; 1ECL.
CC DR TIGR: CC2451; -.
CC DR InterPro: IPR003601; DNAtopi_ATP_bind.
CC DR InterPro: IPR003602; DNAtopi_DNA_bind.
CC DR InterPro: IPR005733; DNA_topi_bact.
CC DR InterPro: IPR000380; DNA_topisomrase.
CC DR InterPro: IPR006171; TopoIim_dom.
CC DR InterPro: IPR006154; TopoIim_sub.
CC DR Pfam: PF01131; TopoIim_bac; 1.
CC DR Pfam: PF01751; TopoIim; 1.
CC DR Pfam: PF01396; zF-C4_TopoIism; 1.
CC DR PRINTS: PR00417; PR018MSRSEI.
CC DR SMART: SM00437; TOP1AC; 1.
CC DR SMART: SM00436; TOP1BC; 1.
CC DR SMART: SM00493; TOP1IM; 1.
CC DR TIGRfams: TIGR01051; topa_bact; 1.
CC DR PROSITE: PS00396; TOPOISOMERASE_I_PROK; 1.
CC DR DNA-binding; Isomerase; Topoisomerase; Complete proteome.
CC KW SEQUENCE 899 AA; 97723 MW; 149SDCEDMD06FA CRC64;
SQ

```

```

OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Ralstoniaceae; Ralstonia.
OX NCBI_TaxID=305;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GMT1000;
RX MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Attiguenave F., Guzy J., Mangenot S.,
RA Arlat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
RA Chandelier M., Choisme N., Claudel-Renard C., Cunha S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Sauvion W., Schlex T.,
RA Signier P., Thebaud P., Whalen M., Winkler P., Levy M.,
RA Weisenbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL Nature 415:497-502(2002).
DR EMBL: AL646071; CAD16500.1; -.
DR InterPro: IPR005819; Histone_H5.
DR PRINTS: PR00624; HISTONEH5.
KW Complete proteome.
KM SEQUENCE 200 AA; 19279 MW; D3831B590510272D CRC64;
SQ

```

Query Match 40.6%; Score 86.5; DB 16; Length 200;
Best Local Similarity 56.5%; Pred. No. 0.24;
Matches 26; Conservative 4; Mismatches 11; Indels 5; Gaps 2;

```

DT 01-OCT-2002 (TReMBLrel. 22, Created)
DT 01-OCT-2002 (TReMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Histone H1.
GN XAC3058.
OS Xanthomonas axonopodis (pv. citri).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=92829;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=306 / ATCC 13902 / XV 101;
RX MEDLINE=22022145; PubMed=12024217;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,
RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA Camarotte G., Camnayan F., Cardozo U., Chamberg F., Clapina L.P.,
RA Ciccarelli R.M.B., Coutinho L.L., Curino-Santos J.R., El-Dorri H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA Fomighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katsuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA Locali B.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidanis U., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA Spinola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezza R.I.D.,
RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA Setubal J.C., Kitajima J.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities.";
RL Nature 417:459-463(2002).
DR EMBL: AE011948; AAM37903.1; -.
DR InterPro: IPR005819; Histone_H5.
DR PRINTS: PR00624; HISTONEH5.
KW Complete proteome.
KM SEQUENCE 155 AA; 16124 MW; FDB53CEB15FD48 CRC64;
SQ

```

Query Match 40.4%; Score 86; DB 16; Length 155;
Best Local Similarity 52.0%; Pred. No. 0.21;

Matches 26; Conservative 4; Mismatches 14; Indels 6; Gaps 2;

Oy 2 KKYAKKAAK-----AKKAYKAAEAKKA-KTEKAAEKAEEAAEAYEA 45
Db 49 KKTAKKATAKKAVKTKATKTAAKKAVKTKAKKATKTAAKKATKTA 98

RESULT 11

O8FP30 ID Q8FP30 PRELIMINARY; PRT; 168 AA.
AC Q8FP30; 01-MAR-2003 (TRENBLrel. 23, Created)
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
DE 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
GN Putative 30S ribosomal protein S16.
OS Corynebacterium efficiens.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacteriaceae; Corynebacteriaceae; Corynebacterium.
OX NCBI_TaxID=152794;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=YS-314 / AJ 12310 / DSM 44549 / JCM 11189;
RA Kawarabayashi Y., Yamazaki J., Hino Y., Kikuchi H., Nakamura Y.,
RA Ikeno K., Suzuki M., Mashima J., Itoh T., Yamagishi A., Nishio Y.,
RA Ueda Y., Sugimoto S.;
RT "The entire genomic sequence of Corynebacterium efficiens YS-314.";
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP005220; BAC18770.1; -.
KW Ribosomal protein; Complete proteome.
SQ SEQUENCE 168 AA; 18228 MW; 443CB68AA2321C01 CRC64;

Query Match 40.4%; Score 86; DB 16; Length 168;
Best Local Similarity 48.9%; Pred. No. 0.22;
Matches 23; Conservative 3; Mismatches 7; Indels 2; Gaps 1;

Oy 7 KAKAEKAKKAYKAAEAKKAYEKAEEKAEEKAEEKA 41
Db 120 EAITEKKKAREEKAEEKA--EKAAEKAEEKAEEKA 152

RESULT 12

O93946 ID O93946 PRELIMINARY; PRT; 212 AA.
AC O93946;
DT 01-MAY-1999 (TRENBLrel. 10, Created)
DT 01-MAY-1999 (TRENBLrel. 10, Last sequence update)
DE 01-OCT-2002 (TRENBLrel. 22, Last annotation update)
GN CTX2p (Fragment).
OS Candida albicans (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX NCBI_TaxID=5476;
RN [1]
RP SEQUENCE FROM N.A.
RA Kaiser B., Kunkel W., Saluz H.P., Munder T.;
RT "Identification of Candida albicans protein domains with
transcriptional activating properties in Saccharomyces cerevisiae.";
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ006637; CA07165.1; -.
FT NON TER 1
SQ SEQUENCE 212 AA; 24231 MW; 10C2122E9554A387 CRC64;

Query Match 40.4%; Score 86; DB 3; Length 212;
Best Local Similarity 48.9%; Pred. No. 0.28;
Matches 23; Conservative 10; Mismatches 12; Indels 2; Gaps 1;

Oy 1 AKKYAKKA--KAEKAKKAYKAAEAKKAYEKAEEKAEEKAEEAYEA 45
Db 70 AKKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAEKAE 116

RESULT 13

O8W044 ID O8W044 PRELIMINARY; PRT; 98 AA.
AC O8W044;
DT 01-MAR-2002 (TRENBLrel. 20, Created)
DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)
DE 01-MAR-2002 (TRENBLrel. 23, Last annotation update)
GN LNP18.
OS Leishmania major.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5664;
RN [1]
RP SEQUENCE FROM N.A.
RA Tzortzis N., Papageorgiou F.T., Tzinia A.K., Soteriadou K.P.;
RT "Identification and characterization of a novel Leishmania gene
encoding for a putative histone H1-like transcription factor.";
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ237814; CAD21431.1; -.
DR InterPro; IPR005819; Histone_H5.
DR PRINTS; PR00624; HISTONEH5.
KW Nuclear protein.
SQ SEQUENCE 98 AA; 9999 MW; 0A4AB93089D6C261 CRC64;

Query Match 39.9%; Score 85; DB 5; Length 98;
Best Local Similarity 55.0%; Pred. No. 0.17;
Matches 22; Conservative 3; Mismatches 15; Indels 0; Gaps 0;

Oy 1 AKKYAKKAEKAEKAYKAAEAKKAYEKAEEKAEEKAEEKA 40
Db 59 AKKPAKKVAKKPAKKAKKPAKKPAKKPAKKPAKKAKKAAKAAK 98

RESULT 14

O8T9R3 ID O8T9R3 PRELIMINARY; PRT; 111 AA.
AC O8T9R3;
DT 01-JUN-2002 (TRENBLrel. 21, Created)
DT 01-JUN-2002 (TRENBLrel. 21, Last sequence update)
DE 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
GN Leishmania infantum.
OS Leishmania infantum.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5671;
RN [1]
RP SEQUENCE FROM N.A.
RA Papageorgiou F., Soteriadou K.;
RT "Identification of a Leishmania infantum gene encoding for an histone
H1-like nuclear protein.";
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF469106; AAL76335.1; -.
KW Nuclear protein.
SQ SEQUENCE 111 AA; 11162 MW; 16168F3B54960B83 CRC64;

Query Match 39.9%; Score 85; DB 5; Length 111;
Best Local Similarity 55.0%; Pred. No. 0.19;
Matches 22; Conservative 3; Mismatches 15; Indels 0; Gaps 0;

Oy 1 AKKYAKKAEKAEKAYKAAEAKKAYEKAEEKAEEKAEEKA 40
Db 72 AKKPAKKVAKKPAKKAKKPAKKPAKKPAKKPAKKAKKAAK 111

RESULT 15

O45370 ID O45370 PRELIMINARY; PRT; 182 AA.
AC O45370;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DE 01-OCT-2002 (TRENBLrel. 22, Last annotation update)
GN Histone H1.
OS Bordetella pertussis.

XX XX

Isteria monocly
Recombinant copoly
Amino acid sequenc
M. tuberculosis hi
Novel human diagn
Mycobacterial hepa
Mycobacterial hepa
Pseudomonas aerug
Mycobacterium bovis
C glutamicum proc
Tumor neoangiogen
Drosophila melano
Drosophila melanog
Histone H1 isoform
Histone H1 isoform
Histone H1 isoform
Histone H1 isoform
Histone H1 isoform
Histone H1 isoform
Human histone H1.5
Human histone H1.5
Human histone H1.1
Human histone H1.1
Human histone H1.1
Human histone H1.1
Human histone H1.1
Drosophila melano
Human ORF proteom
Human ovariin anti
Human polypeptide
Human secreted pte
Human secreted pte
Human ribosomal L
Protein differentia
Human polypeptide

PA (TEVA-) TEVA PHARM USA INC.
 XX
 PI Gad A, Lis D;
 XX
 DR WPI; 2000-317499/27.
 XX
 PT Copolymer 1 related polypeptides used as molecular weight markers for
 PT glatiramer acetate and for treatment and prevention of immune diseases
 XX
 PS Claim 10; Page 14; 72pp; English.
 XX
 CC AAY82571 to AAY82577 represent specifically claimed copolymer molecular
 CC weight TV-marker polypeptides from the present invention. The present
 CC invention describes polypeptides (I) for determining the molecular
 CC weight of a copolymer (CP), which has an identified molecular weight
 CC and an amino acid composition corresponding to the copolymer. The
 CC polypeptides of the invention are used as molecular weight markers for
 CC glatiramer acetate related tetrapolymers. The polypeptides may also be
 CC used for treating and preventing immune diseases in a mammal. Autoimmune
 CC diseases which may be treated include either cell-mediated or
 CC antibody-mediated diseases. Such diseases include arthritic conditions,
 CC sclerolitis, rheumatoid arthritis, osteoarthritis, autoimmune haemolytic
 CC anaemia, autoimmune coprophitis, autoimmune thyroiditis, autoimmune
 CC uveoretinitis, Crohn's disease, chronic immune thrombocytopenia
 CC purpura, colitis, contact sensitivity disease, diabetes mellitus, Graves
 CC disease, Guillain-Barre's syndrome, Hashimoto's disease, idiopathic
 CC myxoedema, myasthenia gravis, psoriasis, pemphigus vulgaris, or systemic
 CC lupus erythematosus. Mediated-mediated diseases which can be treated
 CC include host-versus-graft disease, graft-versus-host disease, and
 CC delayed-type hypersensitivity. The polypeptides of the invention have
 CC defined molecular weights and physical properties which are analogous to
 CC glatiramer acetate molecules, which makes them ideal for use as
 CC molecular weight markers.
 XX
 SQ Sequence 45 AA;
 XX
 Query Match 100.0%; Score 213; DB 21; Length 45;
 Best Local Similarity 100.0%; Pred. No. 1.6e-16;
 Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 AKKYAKKAAEKAKKAYKAAEKAKKAAEKAAEKAAEAAYEA 45
 DB 1 AKKYAKKAAEKAKKAYKAAEKAKKAAEKAKKAAEKAAEAAYEA 45
 XX
 XX
 AC AAY82577;
 XX
 XX 28-JUL-2000 (first entry)
 XX
 DE Copolymer molecular weight TV-marker amino acid sequence SEQ ID NO.7.
 XX
 XX Copolymer; molecular weight marker; TV-marker; immune disease;
 XX glatiramer acetate; autoimmune disease; antiarthritic; neuroprotective;
 XX osteopathic; immunosuppressive; antithyroid; antiinflammatory;
 XX antidiabetic; thyromimetic; haemostatic; antipruritic; dermatological;
 XX antihaemic; immunosuppressive; demyelinating disease; osteoarthritis;
 XX Crohn's disease; chronic immune thrombocytopenia purpura; colitis;
 XX diabetes mellitus; Graves disease; Guillain-Barre's syndrome; psoriasis;
 XX Hashimoto's disease; idiopathic myxoedema; myasthenia gravis;
 XX pemphigus vulgaris; systemic lupus erythematosus.
 XX
 XX Unidentified.
 XX OS
 XX PN WO200018794-A1.
 XX
 XX PD 06-APR-2000.
 XX

PF 24-SEP-1999; 99WO-US22402.
 XX
 PR 25-SEP-1998; 98US-0101693.
 XX
 PA (YEDA) YEDA RES & DEV CO LTD.
 PA (TEVA-) TEVA PHARM USA INC.
 XX
 PI Gad A, Lis D;
 XX
 DR WPI; 2000-317499/27.
 XX
 PT Copolymer 1 related polypeptides used as molecular weight markers for
 PT glatiramer acetate and for treatment and prevention of immune diseases
 XX
 PS Claim 10; Page 14; 72pp; English.
 XX
 CC AAY82571 to AAY82577 represent specifically claimed copolymer molecular
 CC weight TV-marker polypeptides from the present invention. The present
 CC invention describes polypeptides (I) for determining the molecular
 CC weight of a copolymer (CP), which has an identified molecular weight
 CC and an amino acid composition corresponding to the copolymer. The
 CC polypeptides of the invention are used as molecular weight markers for
 CC glatiramer acetate related tetrapolymers. The polypeptides may also be
 CC used for treating and preventing immune diseases in a mammal. Autoimmune
 CC diseases which may be treated include either cell-mediated or
 CC antibody-mediated diseases. Such diseases include arthritic conditions,
 CC sclerolitis, rheumatoid arthritis, osteoarthritis, autoimmune haemolytic
 CC anaemia, autoimmune coprophitis, autoimmune thyroiditis, autoimmune
 CC uveoretinitis, Crohn's disease, chronic immune thrombocytopenia
 CC purpura, colitis, contact sensitivity disease, diabetes mellitus, Graves
 CC disease, Guillain-Barre's syndrome, Hashimoto's disease, idiopathic
 CC myxoedema, myasthenia gravis, psoriasis, pemphigus vulgaris, or systemic
 CC lupus erythematosus. Mediated-mediated diseases which can be treated
 CC include host-versus-graft disease, graft-versus-host disease, and
 CC delayed-type hypersensitivity. The polypeptides of the invention have
 CC defined molecular weights and physical properties which are analogous to
 CC glatiramer acetate molecules, which makes them ideal for use as
 CC molecular weight markers.
 XX
 SQ Sequence 109 AA;
 XX
 Query Match 64.8%; Score 138; DB 21; Length 109;
 Best Local Similarity 76.6%; Pred. No. 5.9e-08;
 Matches 36; Conservative 0; Mismatches 7; Indels 4; Gaps 2;
 QY 1 AKKYAKKAAEKAKKAYKAAEKAKKAAEKAAEKAAEAAYEA 45
 DB 65 AKKYAKKAAEKAKKAYKAAEKAKKAAEKAKKAAEKAAEAAYEA 109
 XX
 XX
 AC AAY82573;
 XX
 XX 28-JUL-2000 (first entry)
 XX
 DE Copolymer molecular weight TV-marker amino acid sequence SEQ ID NO.3.
 XX
 XX Copolymer; molecular weight marker; TV-marker; immune disease;
 XX glatiramer acetate; autoimmune disease; antiarthritic; neuroprotective;
 XX osteopathic; immunosuppressive; antithyroid; antiinflammatory;
 XX antidiabetic; thyromimetic; haemostatic; antipruritic; dermatological;
 XX antihaemic; immunosuppressive; demyelinating disease; osteoarthritis;
 XX Crohn's disease; chronic immune thrombocytopenia purpura; colitis;
 XX diabetes mellitus; Graves disease; Guillain-Barre's syndrome; psoriasis;
 XX Hashimoto's disease; idiopathic myxoedema; myasthenia gravis;
 XX pemphigus vulgaris; systemic lupus erythematosus.
 XX
 XX Unidentified.
 XX OS

XX PN WO200018794-A1.

XX PD 06-APR-2000.

XX PF 24-SEP-1999; 99WO-US22402.

XX PR 25-SEP-1998; 98US-0101693.

XX PA (YEDA) YEDA RES & DEV CO LTD.

XX PA (TEVA-) TEVA PHARM USA INC.

XX P1 Gad A, Lis D;

XX DT WPI; 2000-317499/27.

XX PT Copolymer 1 related polypeptides used as molecular weight markers for

XX PT glatiramer acetate and for treatment and prevention of immune diseases

XX PS Claim 10; Page 14; 72pp; English.

XX AA AAY82571 represents specifically claimed copolymer molecular

XX CC weight TV-marker polypeptides from the present invention. The present

XX CC invention describes polypeptides (I) for determining the molecular

XX CC weight of a copolymer (CP), which has an identified molecular weight

XX CC and an amino acid composition corresponding to the copolymer. The

XX CC polypeptides of the invention are used as molecular weight markers for

XX CC glatiramer acetate related tetrapolymers. The polypeptides may also be

XX CC used for treating and preventing immune diseases in a mammal. Autoimmune

XX CC diseases which may be treated include either cell-mediated or

XX CC antibody-mediated diseases. Such diseases include arthritic conditions,

XX CC scleromyelinating diseases and inflammatory conditions, e.g. multiple

XX CC sclerosis, rheumatoid arthritis, osteoarthritis, autoimmune haemolytic

XX CC anaemia, autoimmune oophoritis, autoimmune thyroiditis, autoimmune

XX CC uveoretinitis, Crohn's disease, chronic immune thrombocytopenia

XX CC purpura, colitis, contact sensitivity disease, diabetes mellitus, Graves

XX CC disease, Guillain-Barre's syndrome, Hashimoto's disease, idiopathic

XX CC myxoedema, myasthenia gravis, psoriasis, pemphigus vulgaris, or systemic

XX CC lupus erythematosus. Mediated diseases which can be treated

XX CC include host-versus-graft disease, graft-versus-host disease, and

XX CC delayed-type hypersensitivity. The polypeptides of the invention have

XX CC defined molecular weights and physical properties which are analogous to

XX CC glatiramer acetate molecules, which makes them ideal for use as

XX CC molecular weight markers.

XX SQ Sequence 56 AA;

XX QY Query Match 63.1%; Score 134.5; DB 21; Length 56;

XX QY Best Local Similarity 68.4%; Pred. No. 7.2e-08;

XX QY Matches 39; Conservative 0; Mismatches 5; Indels 13; Gaps 4;

XX DB 1 AKYAKK-----AKAEK-----KKAYDAEKK--AAKYEKALEKAAEAYEA 45

XX DB 1 AKYAKKKEKAYAKKAEKAAKAAKAAEAYAAAEKKALEAKY-KAEAAKAAAEAYEA 56

XX ID AAY82575 standard; peptide; 77 AA.

XX AC AAY82575;

XX DT 28-JUN-2000 (first entry)

XX DB Copolymer molecular weight TV-marker amino acid sequence SEQ ID NO:5.

XX KM Copolymer; molecular weight marker; TV-marker; immune disease;

XX KM glatiramer acetate; autoimmune disease; antiarthritis; neuroprotective;

XX KM osteoprotective; immunosuppressive; antithyroid; antiinflammatory;

XX KM antidiabetic; thymostimetic; haemostatic; antiproliferative; dermatological;

XX KM antianemic; immunosuppressive; demyelinating disease; osteoarthritis;

XX KM inflammatory condition; multiple sclerosis; rheumatoid arthritis;

XX KM Crohn's disease; chronic immune thrombocytopenia purpura; colitis;

```

XX diabetes mellitus; Graves disease; Guillain-Barre's syndrome; psoriasis;
XX Hashimoto's disease; idiopathic myxoedema; myasthenia gravis;
XX pemphigus vulgaris; systemic lupus erythematosus.
XX
XX unidentified.
XX
XX WO200018794-A1.
XX
XX 06-APR-2000.
XX
XX 24-SEP-1999; 99WO-US22402.
XX
XX 25-SEP-1998; 98US-0101693.
XX
XX (YEDA ) YEDA RES & DEV CO LTD.
XX (TEVA-) TEVA PHARM USA INC.
XX
XX Gad A, Iis D;
XX
XX WPI; 2000-317499/27.
XX
XX Copolymer 1 related polypeptides used as molecular weight markers for
XX glatiramer acetate and for treatment and prevention of immune diseases
XX
XX Claim 10; Page 14; 72pp; English.
XX
XX AAy82571 to AAy82577 represent specifically claimed copolymer molecular
XX weight TV-marker polypeptides from the present invention. The present
XX invention describes polypeptides (I) for determining the molecular
XX weight of a copolymer (CP), which has an identified molecular weight
XX and an amino acid composition corresponding to the copolymer. The
XX polypeptides of the invention are used as molecular weight markers for
XX glatiramer acetate related tetrapolymers. The polypeptides may also be
XX used for treating and preventing immune diseases in a mammal. Autoimmune
XX diseases which may be treated include either cell-mediated or
XX antibody-mediated diseases. Such diseases include arthritic conditions,
XX demyelinating diseases and inflammatory conditions, e.g. multiple
XX sclerosis, rheumatoid arthritis, osteoarthritis, autoimmune haemolytic
XX anaemia, autoimmune oophoritis, autoimmune thyroiditis, autoimmune
XX uveoretinitis, Crohn's disease, chronic immune thrombocytopenia
XX purpura, colitis, contact sensitivity disease, diabetes mellitus, Graves
XX disease, Guillain-Barre's syndrome, Hashimoto's disease, idiopathic
XX myxoedema, myasthenia gravis, psoriasis, pemphigus vulgaris, or systemic
XX lupus erythematosus. Mediated-mediated diseases which can be treated
XX include host-versus-graft disease, graft-versus-host disease, and
XX delayed-type hypersensitivity. The polypeptides of the invention have
XX defined molecular weights and physical properties which are analogous to
XX glatiramer acetate molecules, which makes them ideal for use as
XX molecular weight markers.
XX
XX SQ Sequence 77 AA;
XX
XX Query Match 61.5%; Score 131; DB 21; Length 77;
XX Best Local Similarity 50.6%; Pred. No. 2.4e-07;
XX Matches 39; Conservative 0; Mismatches 6; Indels 32; Gaps 3;
XX
XX QY 1 AKKYAKK-----AAAEKA-----KKAYKAAEAKKAYE----- 29
XX ||||| ||||| |||||
XX DB 1 AKKYAKKEKAYAKKAEKAKKAEKAYKAAAEAKKKKAAKAEAKKAYAKKAAEKEKAYAAAEAK 60
XX ||||| ||||| |||||
XX QY 30 -KAAAEKAAAEAAVEA 45
XX ||||| ||||| |||||
XX DB 61 YKAEAAKAAAEAAVEA 77
XX
XX RESULT 5
XX AAy82576
XX ID AAy82576 standard; peptide; 86 AA.
XX
XX AC AAy82576;
XX
XX 28-JUL-2000 (first entry)
XX

```

XX	Copolymer molecular weight TV-marker amino acid sequence SEQ ID NO:6.
KW	Copolymer; molecular weight marker; TV-marker; immune disease;
KW	glucosamine; acetate; autoimmune disease; antiarthritic; neuroprotective;
KW	osteopathic; immunosuppressive; antithyroid; antiinflammatory;
KW	antidiabetic; thyromimetic; haemostatic; antipsoriatic; dermatological;
KW	antihaemic; immunosuppressive; demyelinating disease; osteoarthritis;
KW	inflammatory condition; multiple sclerosis; rheumatoid arthritis;
KW	Crohn's disease; chronic immune thrombocytopenia purpura; colitis;
KW	diabetes mellitus; Graves disease; Guillain-Barre's syndrome; psoriasis;
KW	Hashimoto's disease; idiopathic myxoedema; myasthenia gravis;
KW	pemphigus vulgaris; systemic lupus erythematosus.
XX	
OS	Unidentified.
XX	
XX	WO200018794-A1.
PD	
XX	06-APR-2000.
XX	
XX	24-SEP-1999; 99WO-US22402.
XX	
XX	25-SEP-1998; 98US-0101693.
XX	
PA	(YEDA) YEDA RES & DEV CO LTD.
PA	(TEVA-) TEVA PHARM USA INC.
XX	
PI	Gad A, Lis D;
XX	
DR	WPI; 2000-317499/27.
PT	
XX	Copolymer 1 related polypeptides used as molecular weight markers for
XX	glutaram acetate and for treatment and prevention of immune diseases
XX	Claim 10; Page 14; 72pp; English.

Query Match	Similarity	Score	DB	Length
Best Local	Similarity 45.3%	Pred. No. 8.3e-07		
Matches	39; Conservative	0; Mismatches	6; Indels	41; Gaps

Sequence	86 AA;
1 AKTKAKK-----AKAEKA-----KATYKAAEKKAATK-----	29
1 AKTKAKKKAAYAKKAEKKAAKKAKEAAKAYKAAAEKKAATKAEKTKYAAKAEKKEKAAAEAK	60
30 -----KAAAEKAAKKAAYEA	45
61 YKAEAKKAYKAAEAKAAAEKKAAYEA	86

XX	AAV82574	standard; peptide; 66 AA.
XX	AAV82574;	
XX	28-JUL-2000	(first entry)
XX	Copolymer	molecular weight TV-marker amino acid sequence SEQ ID NO:4.
XX	Copolymer;	molecular weight marker; TV-marker; immune disease;
XX	glucitamer;	acetate; autoimmune disease; antiarthritic; neuroprotective;
XX	osteopathic;	immunosuppressive; antithyroid; antiinflammatory;
XX	antidiabetic;	thyromimetic; haemostatic; antiproliferative; dermatological;
XX	antianaemic;	immunosuppressive; demyelinating disease; osteoarthritis;
XX	inflammatory condition;	multiple sclerosis; rheumatoid arthritis;
XX	Crohn's disease;	chronic immune thrombocytopenia purpura; colitis;
XX	diabetes mellitus;	Graue's disease; Guillain-Barre's syndrome; psoriasis;
XX	Hashimoto's disease;	idiopathic myxoedema; myasthenia gravis;
XX	periphagus vulgaris;	systemic lupus erythematosus.
XX	Unidentified.	
XX	WO200018794-A1.	
XX	06-APR-2000.	
XX	24-SEP-1999;	99WO-US22402.
XX	25-SEP-1998;	98US-0101693.
XX	(YEDA)	YEDA RES & DEV CO LTD.
XX	(TEVA-)	TEVA PHARM USA INC.
XX	Gad A,	Lis D;
XX	WPI;	2000-317499/27.
XX	Copolymer 1	related polypeptides used as molecular weight markers for
XX	glucitamer	acetate and for treatment and prevention of immune diseases
XX	Claim 10;	Page 14; 72pp; English.

AA82577 represents specifically claimed copolymer molecular weight TV-marker polypeptides from the present invention. The present invention describes polypeptides (I) for determining the molecular weight of a copolymer (CP), which has an identified molecular weight and an amino acid composition corresponding to the copolymer. The polypeptides of the invention are used as molecular weight markers for glutarimer acetate related tetrapolymers. The polypeptides may also be used for treating and preventing immune diseases in a mammal. Autoimmune diseases which may be treated include either cell-mediated or antibody-mediated diseases. Such diseases include arthritic conditions, demyelinating diseases and inflammatory conditions, e.g. multiple sclerosis, rheumatoid arthritis, osteoarthritis, autoimmune haemolytic anaemia, autoimmune uveoretinitis, autoimmune thyroiditis, autoimmune uveoretinitis, Crohn's disease, chronic immune thrombocytopenia, purpura, colitis, contact sensitivity disease, diabetes mellitus, Graves disease, Guillain-Barre's syndrome, Hashimoto's disease, idiopathic myxoedema, myasthenia gravis, psoriasis, pemphigus vulgaris, or systemic lupus erythematosus. Mediated-mediated diseases which can be treated include host-versus-graft disease, graft-versus-host disease, and delayed-type hypersensitivity. The polypeptides of the invention have defined molecular weights and physical properties which are analogous to glutarimer acetate molecules, which makes them ideal for use as molecular weight markers.

KM vitamin B12; bacterial infection; disease.
 XX Listeria monocytogenes.
 OS
 XX WO200177335-A2.
 XX
 PD 18-OCT-2001.
 XX
 PF 11-APR-2001; 2001WO-FR01118.
 XX
 PR 11-APR-2000; 2000FR-0004629.
 XX
 PA (INSP) INST PASTEUR.
 XX Buchrieser C, Frangeul L, Couve E, Rusniok C, Eshti H, Dehoux P,
 PI Duesbrugel O, Chetoui F, Nedjati H, Glaser P, Kunst F, Cossart P,
 PI Daniels J, Goebel W, Kreft J, Kuhn M, Ng B, Vazquez-Boland JA,
 PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A,
 PI Chakraborty T, Dommann E, Hain T, Berche P, Charbit A, Durant L,
 PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N,
 PI Madueno E, De Pablo B, Weiland U, Kaerst U, Entian K, Hauf J,
 PI Rose M, Voss H,
 DR WPI; 2002-010914/01.
 XX
 PT Genomic sequence for Listeria monocytogenes, useful e.g. for treatment
 PT and prevention of Listeria and related bacterial infections, and
 PT related polypeptides
 PS Claim 6; SEQ ID No 1828; 192pp; French.
 XX
 CC The present invention relates to the genome sequence of Listeria
 CC monocytogenes EGD-e (see ABA03041). The genome sequence and fragments of
 CC it are useful for selecting probes and primers for detecting genes in L.
 CC monocytogenes and related organisms, and for studying genetic
 CC polymorphisms and other genomes. The present sequence is a protein
 CC encoded by the genome sequence of the present invention. Proteins
 CC expressed from the genome sequence are useful for raising specific
 CC antibodies, identification of L. monocytogenes and related organisms, and
 CC for biosynthesis and biodegradation, especially biosynthesis of Vitamin
 CC B12. The genome sequence and proteins encoded by it are also useful for
 CC selecting compounds that regulate gene expression and cell replication
 CC and modulate L. monocytogenes-related diseases. In addition, the genome
 CC sequence and proteins encoded by it are useful in pharmaceutical and
 CC vaccines compositions for the treatment or prevention of infections by L.
 CC monocytogenes and related organisms.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 CC
 SQ Sequence 239 AA;
 Qy Query Match 44.4%; Score 94.5; DB 23; Length 239;
 Best Local Similarity 56.2%; Pred. No. 0.0072;
 Matches 27; Conservative 7; Mismatches 9; Indels 5; Gaps 2;
 Db 2 KKYAKKAAEKAKK--AYKAAEK--AATYKAAKAAKAAKAAAYE 44
 124 KAAAEKAEADKKQEDAVKAAKAEQEAEEKAAADKAAAEKAAAE 171
 RESULT 11
 ID AAR06445 standard; protein; 154 AA.
 XX AAR06445;
 XX
 DT 25-MAR-2003 (updated)
 DT 03-JAN-1991 (first entry)
 XX Recombinant copolymer 1-77, myelin basic protein analogue.
 DE Recombinant copolymer 1; COP-1-77; myelin basic protein; MBP;
 KW

KM immunological activity; autoimmune encephalomyelitis;
 KW multiple sclerosis;
 XX
 OS Synthetic.
 XX
 PN EP383620-A.
 XX
 PD 22-AUG-1990.
 XX
 PF 16-FEB-1990; 90EP-0301700.
 XX
 PR 07-FEB-1990; 90US-0473845.
 PR 17-FEB-1989; 89US-0312541.
 XX
 PA (REPK) REPLIGEN CORP.
 XX
 PI Cook KS;
 DR WPI; 1990-255848/34.
 DR N-P8DB; AAQ05664.
 XX
 PT Producing genes encoding random polymers of aminoacid(s) - for
 PT producing recombinant polypeptide(s) with biological and/or
 PT immunological activity
 PS Disclosure; Fig 11; 25pp; English.
 XX
 CC To improve the expression of rCOP-1 polypeptides in E. coli, genes
 CC coding for rCOP-1-77 were subcloned from pREV 2.1 to pBG3-2delAN
 CC (deposit: 20-NOV-1984 US4691009, NRR1 B-15910), a plasmid used to
 CC express Protein A. The resulting plasmids encode fusion proteins
 CC consisting of beta-glucuronidase, Protein A, and rCOP-1
 CC A methionine residue occurs between the Protein A and rCOP-1
 CC sequences, originating from the 5' linker sequence, in order that
 CC the COP-1 polypeptide may be cleaved from the fusion protein.
 CC rCOP-1-77 contains oligonucleotide duplexes encoding the following
 CC segments: YKK, EAK, KKK, AAK, and AAA. The N-terminal alanine residue
 CC is left behind following CNBr cleavage of the fusion protein.
 CC The product prevents or arrests experimental autoimmune
 CC encephalomyelitis. They are used to prevent, arrest or control a
 CC demyelinating disorder, e.g. multiple sclerosis. They may also
 CC be used as additives to hair care products to confer beneficial
 CC effects on damaged hair or as supplements for diets deficient in
 CC certain amino acids.
 CC See also AAQ05665.
 CC (Updated on 25-MAR-2003 to correct PA field.)
 CC
 SQ Sequence 154 AA;
 Qy Query Match 44.1%; Score 94; DB 11; Length 154;
 Best Local Similarity 60.5%; Pred. No. 0.0052;
 Matches 26; Conservative 2; Mismatches 9; Indels 6; Gaps 2;
 Db 2 KKYAKKAAEKAKKAYKAAEKAAKAAKAAKAAAYE 44
 105 KKYAKKAAEKAKKAYK--KAKAEKKA-----KAAAEKAAKAAKAAEYK 141
 RESULT 12
 ID AAY14928 standard; protein; 223 AA.
 XX AAY14928;
 XX
 DT 25-OCT-1999 (first entry)
 DT
 XX Amino acid sequence of M. vaccae antigen GV-45.
 KW Mycobacterium vaccae protein; antigen; T cell activation; cytokine;
 KW dendritic cell maturation; infectious disease; immune disorder; cancer;
 KW respiratory system; mycobacterial infection; allergy; tuberculosis;
 KW leprosy; sarcoidosis; lung cancer; asthma; skin disorder; psoriasis;
 KW dermatitis; eczema; alopecia areata; skin cancer; basal carcinoma;
 KW


```

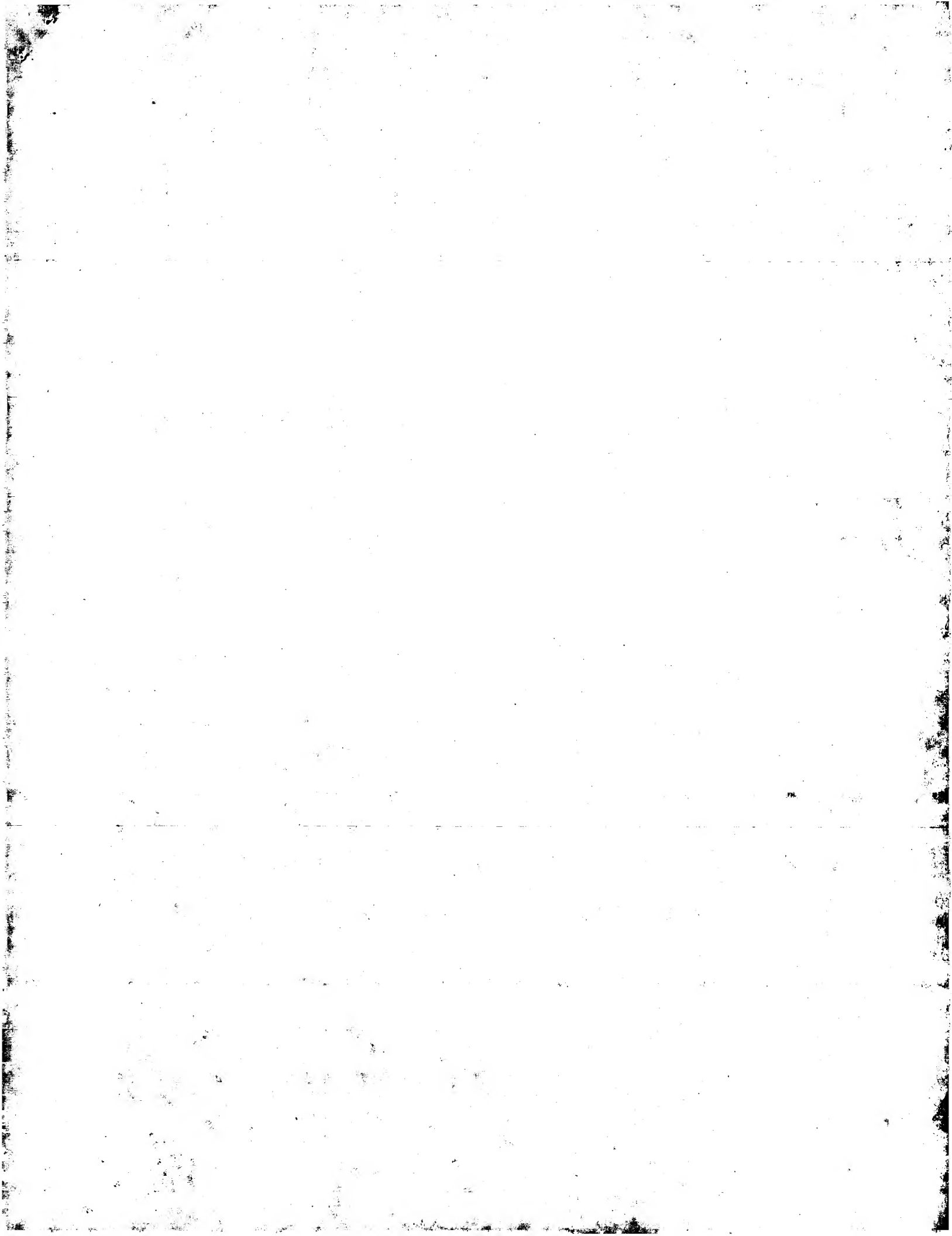
PD XX      16-SBP-1999.
PF XX      12-MAR-1999;    99NO-US05492.
PR XX      12-MAR-1998;    98US-0041889.
PA XX      (REGC ) UNIV CALIFORNIA.
PI XX      Braun J, Cohavy O;
XX DR      WPI; 1999-551215/46.
PT XX      Use of histone H1, porin or Bacteroides antigens as targets for the
PS XX      diagnosis, prevention and treatment of ulcerative colitis -
        Claim 2, Fig 11, 134pp; English.
XX PS      The invention provides a method for the diagnosis, prevention and
CC CC      treatment of ulcerative colitis (UC) using histone H1-like antigen, a
CC CC      porin antigen or a Bacteroides antigen as a target antigen. The novel
CC CC      method of diagnosing UC in a subject suspected of having inflammatory
CC CC      bowel disease (IBD) comprises: (1) obtaining a sample from the subject;
CC CC      (2) contacting the sample with a histone H1-like antigen, or perinuclear
CC CC      anti-neutrophil cytoplasmic antibody (panCA)-reactive fragment, to form a
CC CC      complex of the histone H1-like antigen, or the panCA-reactive fragment,
CC CC      and antibody to the histone H1-like antigen; and (3) detecting the
CC CC      presence or absence of the complex; where the presence of the complex
CC CC      indicates that the subject has UC. The panCA-reactive histone H1-like
CC CC      antigen, porin antigen and Bacteroides antigen are useful in the
CC CC      diagnosis, prevention and treatment of UC. The methods can also be used
CC CC      for identifying agents useful for treating UC. The present sequence
CC CC      represents a M. tuberculosis histone H1-like antigen.
XX SQ      Sequence      214 AA;

Query Match          42.0%; Score 89.5; DB 20; Length 214;
Best Local Similarity 55.6%; Pred. No. 0.023;
Matches   25; Conservative   4; Mismatches   13; Indels   3; Gaps   1.

QY      1 AKKAKKAAGAEEAKKAYKAAEAKKATYEKAAAEKAAKEAAYEA 45
       ||| ||||| :||| :||| :||| :||| :||| :||| :|||
DB      111 AKGVAKGAPAKKATTKAAKKAAATKAPA---RKAATKAPAKKAAATKA 152

RESULT 14
ID ID AAY57353 standard; Protein; 214 AA.
XX AC AAY57353;
XX DT 13-JUN-2000 (first entry)
XX DE M. tuberculosis histone H1-like protein, 214.
XX KW Ulcerative colitis; inflammatory bowel disease; porin antigen; MAAb;
RW PANCA; perinuclear anti-neutrophil cytoplasmic antibody; 214 protein;
XX histone H1; isoform.
XX OS Mycobacterium tuberculosis.
XX FN US6033864-A.
XX PD 07-MAR-2000.
XX PF 12-MAR-1998;    98US-0041889.
XX PR 12-APR-1996;    96US-0057846.
XX PR 11-APR-1997;    97US-0837058.
XX PA (REGC ) UNIV CALIFORNIA.
XX PI Cohavy O, Braun J;

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 21, 2004, 09:01:02 ; Search time 16.3636 Seconds
(without alignments)
562.323 Million cell updates/sec

Title: US-09-816-989A-2

Perfect score: 1 AKKYAKKAKKAKKAKKAYKA.....AKTEKAAAEKAAKEAAVEA 45

Sequence:

Scoring table:

Gapop 10.0 , Gapext 0.5

Searched: 762491 seqs, 204481190 residues

Total number of hits satisfying chosen parameters: 762491

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/FC1_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_NEW_PUB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_NEW_PUB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	213	100.0	45	US-09-816-989A-2	Sequence 2, Appli
2	138	64.8	109	US-09-816-989A-7	Sequence 7, Appli
3	134.5	61.1	56	US-09-816-989A-3	Sequence 3, Appli
4	131	61.5	77	US-09-816-989A-5	Sequence 5, Appli
5	126.5	59.4	86	US-09-816-989A-6	Sequence 6, Appli
6	120.5	56.6	66	US-09-816-989A-4	Sequence 4, Appli
7	108	50.7	35	US-09-816-989A-1	Sequence 1, Appli
8	91	42.7	223	US-10-205-979-52	Sequence 52, Appli
9	91	42.7	223	US-10-051-643-201	Sequence 201, App
10	89.5	40.4	214	US-10-229-567-27	Sequence 27, Appli
11	86	40.4	347	US-10-127-032-120	Sequence 120, App
12	85	39.9	376	US-10-156-761-9889	Sequence 9889, Ap
13	82	38.5	165	US-09-738-626-5751	Sequence 5751, Ap
14	78	36.6	272	US-10-156-761-12370	Sequence 12370, A
15	78	36.6	309	US-09-820-843A-24	Sequence 24, Appli

16	77.5	36.4	227	US-10-156-761-10923	Sequence 10923, A
17	77	36.2	103	US-10-229-567-39	Sequence 39, Appli
18	77	36.2	116	US-10-229-567-38	Sequence 38, Appli
19	77	36.2	158	US-10-229-567-40	Sequence 40, Appli
20	77	36.2	222	US-10-229-567-32	Sequence 32, Appli
21	77	36.2	226	US-10-229-567-32	Sequence 32, Appli
22	75.5	35.4	201	US-10-264-049-2597	Sequence 2597, Ap
23	75.5	35.4	220	US-09-923-104-2	Sequence 2, Appli
24	75.5	35.4	255	US-09-925-102-559	Sequence 559, App
25	74.5	35.0	1002	US-10-369-493-1378	Sequence 1378, Ap
26	74	34.7	352	US-09-820-843A-23	Sequence 23, Appli
27	74	34.7	356	US-09-820-843A-27	Sequence 27, Appli
28	74	34.7	942	US-10-156-761-12155	Sequence 12155, A
29	73	34.3	324	US-10-262-209-1	Sequence 1, Appli
30	72.5	34.0	372	US-09-820-843A-8	Sequence 8, Appli
31	71	33.3	110	US-10-262-209-2	Sequence 2, Appli
32	71	33.3	218	US-10-229-567-4	Sequence 4, Appli
33	71	33.3	228	US-10-156-761-9425	Sequence 9425, Ap
34	71	33.3	539	US-10-369-493-17058	Sequence 17058, A
35	70.5	33.1	548	US-10-128-714-3127	Sequence 1127, Ap
36	70.5	33.1	555	US-10-128-714-8127	Sequence 8127, Ap
37	70.5	33.1	619	US-09-882-774-1	Sequence 1, Appli
38	70	32.9	55	US-10-229-567-41	Sequence 41, Appli
39	70	32.9	347	US-10-156-761-11652	Sequence 11652, A
40	70	32.9	641	US-09-769-787-184	Sequence 184, App
41	70	32.9	744	US-09-815-242-13187	Sequence 13187, A
42	70	32.9	1156	US-10-032-585-7350	Sequence 7350, Ap
43	69.5	32.6	1130	US-10-369-493-6751	Sequence 6751, Ap
44	69.5	32.6	12	US-10-156-761-13000	Sequence 13000, A
45	69.5	32.6	1387		

ALIGNMENTS

RESULT 1
US-09-816-989A-2
Sequence 2, Application US/09816989A
Patent No. US20020115103A1
GENERAL INFORMATION:
APPLICANT: Gad, Alexander
TITLE OF INVENTION: COPOLYMER 1 RELATED POLYPEPTIDES FOR USE AS MOLECULAR WEIGHT MARK
FILE REFERENCE: 2609/60807-A-PCT-US
CURRENT FILING DATE: 2001-03-23
PRIOR APPLICATION NUMBER: 60/101,693
PRIOR FILING DATE: 1998-09-25
PRIOR APPLICATION NUMBER: PCT/US99/22402
PRIOR FILING DATE: 1999-09-24
NUMBER OF SEQ ID NOS: 7
SOFTWARE: PatentIn version 3.1
SEQ ID NO 2
LENGTH: 45
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Synthetic Peptide
US-09-816-989A-2
Query Match 100.0%; Score 213; DB 10; Length 45;
Best Local Similarity 100.0%; Pred. No. 1,4e-16;
Matches 45; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Cy 1 AKKYAKKAKKAKKAKKAYKA.....AKTEKAAAEKAAKEAAVEA 45
Db 1 AKKYAKKAKKAKKAKKAYKA.....AKTEKAAAEKAAKEAAVEA 45
RESULT 2
US-09-816-989A-7
Sequence 7, Application US/09816989A

```
/ Patent No. US20020115103A1
/ GENERAL INFORMATION:
/ APPLICANT: Lis, Doris
/ APPLICANT: Gad, Alexander
/ TITLE OF INVENTION: COPOLYMER 1 RELATED POLYPEPTIDES FOR USE AS MOLECULAR WEIGHT MARK
/ FILE REFERENCE: 2609/60807-A-PCT-US
/ CURRENT APPLICATION NUMBER: US/09/816,989A
/ PRIOR FILING DATE: 2001-03-23
/ PRIOR APPLICATION NUMBER: 60/101,693
/ PRIOR FILING DATE: 1998-09-25
/ PRIOR APPLICATION NUMBER: PCT/US99/22402
/ NUMBER OF SEQ ID NOS: 7
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 7
/ LENGTH: 109
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Description of Artificial Sequence: Synthetic Peptide
US-09-816-989A-7

Query Match          64.8%; Score 138; DB 10; Length 109;
Best Local Similarity 76.6%; Pred. No. 4.8e-08;
Matches 36; Conservative 0; Mismatches 7; Indels 4; Gaps 2;

Qy 1 AKKYAKK-----AKAEKA-----KKAYKAAEKKAAYE 45
    |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 65 AKKYAKKAE--KKAYKAAEKKAAYKAAYKAAYKAAYKAAYEA 109

RESULT 3
US-09-816-989A-3
/ Sequence 3, Application US/09816989A
/ Patent No. US20020115103A1
/ GENERAL INFORMATION:
/ APPLICANT: Lis, Doris
/ APPLICANT: Gad, Alexander
/ TITLE OF INVENTION: COPOLYMER 1 RELATED POLYPEPTIDES FOR USE AS MOLECULAR WEIGHT MARK
/ FILE REFERENCE: 2609/60807-A-PCT-US
/ CURRENT APPLICATION NUMBER: US/09/816,989A
/ PRIOR FILING DATE: 2001-03-23
/ PRIOR APPLICATION NUMBER: 60/101,693
/ PRIOR FILING DATE: 1998-09-25
/ PRIOR APPLICATION NUMBER: PCT/US99/22402
/ NUMBER OF SEQ ID NOS: 7
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 3
/ LENGTH: 56
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Description of Artificial Sequence: Synthetic Peptide
US-09-816-989A-3

Query Match          63.1%; Score 134.5; DB 10; Length 56;
Best Local Similarity 68.4%; Pred. No. 5.6e-08;
Matches 39; Conservative 0; Mismatches 5; Indels 13; Gaps 4;

Qy 1 AKKYAKK-----AKAEKA-----KKAYKAAEKKAAYE 45
    |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 1 AKKYAKKKAAYKAAYKAAYKAAYKAAYKAAYKAAYKAAYEA 56

RESULT 4
US-09-816-989A-5
/ Sequence 5, Application US/09816989A
/ Patent No. US20020115103A1
/ GENERAL INFORMATION:
/ APPLICANT: Gad, Alexander
```

```
/ APPLICANT: Lis, Doris
/ TITLE OF INVENTION: COPOLYMER 1 RELATED POLYPEPTIDES FOR USE AS MOLECULAR WEIGHT MARK
/ FILE REFERENCE: 2609/60807-A-PCT-US
/ CURRENT APPLICATION NUMBER: US/09/816,989A
/ PRIOR FILING DATE: 2001-03-23
/ PRIOR APPLICATION NUMBER: 60/101,693
/ PRIOR FILING DATE: 1998-09-25
/ PRIOR APPLICATION NUMBER: PCT/US99/22402
/ NUMBER OF SEQ ID NOS: 7
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 5
/ LENGTH: 77
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Description of Artificial Sequence: Synthetic Peptide
US-09-816-989A-5

Query Match          61.5%; Score 131; DB 10; Length 77;
Best Local Similarity 50.6%; Pred. No. 1.9e-07;
Matches 39; Conservative 0; Mismatches 6; Indels 32; Gaps 3;

Qy 1 AKKYAKK-----AKAEKA-----KKAYKAAEKKAAYE 29
    |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 1 AKKYAKKKAAYKAAYKAAYKAAYKAAYKAAYKAAYKAAYEA 60

Qy 30 -KAAEKAAYKAAYEA 45
    |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 61 YKAEAAYKAAYKAAYKAAYKAAYKAAYKAAYKAAYEA 77

RESULT 5
US-09-816-989A-6
/ Sequence 6, Application US/09816989A
/ Patent No. US20020115103A1
/ GENERAL INFORMATION:
/ APPLICANT: Lis, Doris
/ APPLICANT: Gad, Alexander
/ TITLE OF INVENTION: COPOLYMER 1 RELATED POLYPEPTIDES FOR USE AS MOLECULAR WEIGHT MARK
/ FILE REFERENCE: 2609/60807-A-PCT-US
/ CURRENT APPLICATION NUMBER: US/09/816,989A
/ PRIOR FILING DATE: 2001-03-23
/ PRIOR APPLICATION NUMBER: 60/101,693
/ PRIOR FILING DATE: 1998-09-25
/ PRIOR APPLICATION NUMBER: PCT/US99/22402
/ NUMBER OF SEQ ID NOS: 7
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 6
/ LENGTH: 86
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Description of Artificial Sequence: Synthetic Peptide
US-09-816-989A-6

Query Match          59.4%; Score 126.5; DB 10; Length 86;
Best Local Similarity 45.3%; Pred. No. 6.5e-07;
Matches 39; Conservative 0; Mismatches 6; Indels 41; Gaps 3;

Qy 1 AKKYAKK-----AKAEKA-----KKAYKAAEKKAAYE 29
    |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 1 AKKYAKKKAAYKAAYKAAYKAAYKAAYKAAYKAAYKAAYEA 60

Qy 30 -----KAAEKAAYKAAYEA 45
    |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 61 YKAEAAYKAAYKAAYKAAYKAAYKAAYKAAYKAAYEA 86

RESULT 6
```

```

US-09-816-989A-4
; Sequence 4, Application US/09816989A
; Patent No. US20020115103A1
; GENERAL INFORMATION:
; APPLICANT: Gad, Alexander
; APPLICANT: Lieg, Doris
; TITLE OF INVENTION: COPOLYMER 1 RELATED POLYPEPTIDES FOR USE AS MOLECULAR WEIGHT MARK
; FILE REFERENCE: 2609/60807-A-PCT-US
; CURRENT APPLICATION NUMBER: US/09/816,989A
; PRIOR FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: 60/101,693
; PRIOR FILING DATE: 1998-09-25
; PRIOR APPLICATION NUMBER: PCT/US99/22402
; PRIOR FILING DATE: 1999-09-24
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 4
; LENGTH: 66
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic Peptide
US-09-816-989A-4

Query Match          56.6%; Score 120.5; DB 10; Length 66;
Best Local Similarity 71.1%; Pred. No. 2.2e-06;
Matches 32; Conservative 0; Mismatches 6; Indels 7; Gaps 2;

Cy 1 AKKYAKKAKKAKKAYKAAEKKAAYEKKAABKAAKAAEAYEA 45
Db 29 AKKYAKKAKKAE-KKETAABERK-----YKAAKAAKAAKAAEAYEA 66

RESULT 7
US-09-816-989A-1
; Sequence 1, Application US/09816989A
; Patent No. US20020115103A1
; GENERAL INFORMATION:
; APPLICANT: Gad, Alexander
; APPLICANT: Lieg, Doris
; TITLE OF INVENTION: COPOLYMER 1 RELATED POLYPEPTIDES FOR USE AS MOLECULAR WEIGHT MARK
; FILE REFERENCE: 2609/60807-A-PCT-US
; CURRENT APPLICATION NUMBER: US/09/816,989A
; PRIOR FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: 60/101,693
; PRIOR FILING DATE: 1998-09-25
; PRIOR APPLICATION NUMBER: PCT/US99/22402
; PRIOR FILING DATE: 1999-09-24
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 1
; LENGTH: 35
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic Peptide
US-09-816-989A-1

Query Match          50.7%; Score 108; DB 10; Length 35;
Best Local Similarity 64.4%; Pred. No. 2.5e-05;
Matches 29; Conservative 1; Mismatches 5; Indels 10; Gaps 2;

Cy 1 AKKYAKKAKKAKKAYKAAEKKAAYEKKAABKAAKAAEAYEA 45
Db 1 AKKYAKKAKKAE-KKAYK-----KAAKAAKAAKAAKAAEAYEA 35

RESULT 8
US-10-205-979-52
; Sequence 52, Application US/10205979
; Publication No. US20030147861A1

```

```

; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Tan, Paul L. J.
; APPLICANT: Abernethy, Nevin
; TITLE OF INVENTION: Compounds and Methods for the Modulation
; FILE REFERENCE: 11000.1063U
; CURRENT APPLICATION NUMBER: US/10/205,979
; PRIOR FILING DATE: 2002-07-25
; PRIOR APPLICATION NUMBER: 60/308,446
; PRIOR FILING DATE: 2001-07-26
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 52
; LENGTH: 223
; TYPE: PRT
; ORGANISM: Mycobacterium vaccae
US-10-205-979-52

Query Match          42.7%; Score 91; DB 12; Length 223;
Best Local Similarity 59.6%; Pred. No. 0.013;
Matches 28; Conservative 5; Mismatches 12; Indels 2; Gaps 2;

Cy 1 AKKYA-KKAKAKKAKKAYKAAEKKAAYEKKA-ABKAAKAAEAYEA 45
Db 137 AKKAATTAAPAKKATTAAKKAAPAKKATPAKKAAPAKKATPAKKAATKA 183

RESULT 9
US-10-051-643-201
; Sequence 201, Application US/10051643
; Publication No. US20020197265A1
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Tan, Paul L. J.
; TITLE OF INVENTION: Methods and Compounds for the Treatment
; TITLE OF INVENTION: of Immunologically-Mediated Diseases of the Respiratory
; FILE REFERENCE: 11000.1008C2
; CURRENT APPLICATION NUMBER: US/10/051,643
; PRIOR FILING DATE: 2002-01-18
; PRIOR APPLICATION NUMBER: US09/156,181
; PRIOR FILING DATE: 1998-09-17
; PRIOR APPLICATION NUMBER: US 08/996,624
; PRIOR FILING DATE: 1997-12-23
; NUMBER OF SEQ ID NOS: 208
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 201
; LENGTH: 223
; TYPE: PRT
; ORGANISM: Mycobacterium vaccae
US-10-051-643-201

Query Match          42.7%; Score 91; DB 14; Length 223;
Best Local Similarity 59.6%; Pred. No. 0.013;
Matches 28; Conservative 5; Mismatches 12; Indels 2; Gaps 2;

Cy 1 AKKYA-KKAKAKKAKKAYKAAEKKAAYEKKA-ABKAAKAAEAYEA 45
Db 137 AKKAATTAAPAKKATTAAKKAAPAKKATPAKKAAPAKKATPAKKAATKA 183

RESULT 10
US-10-229-567-27
; Sequence 27, Application US/10229567
; Publication No. US20030092080A1
; GENERAL INFORMATION:
; APPLICANT: Braun, Jonathan
; APPLICANT: Chavvy, Ofer
; TITLE OF INVENTION: Diagnosis, Prevention and Treatment of
; ulcerative Colitis, and Clinical Subtypes Thereof, Using
; Microbial UC PANCA antigens
; NUMBER OF SEQUENCES: 41

```


TYPE: PRT
ORGANISM: Corynebacterium glutamicum
US-09-738-626-5751

Query Match 38.5%; Score 82; DB 10; Length 165;
Best Local Similarity 62.9%; Pred. No. 0.087;
Matches 22; Conservative 4; Mismatches 7; Indels 2; Gaps 1;

QY 7 KAKAEKAKKAYKAAAEKAKYKAEKAAAEKAAKEA 41
DB 120 EAITEKKKKAREDEKAEKA--EKAAAEKAAAS 152

RESULT 14
US-10-156-761-12370
Sequence 12370, Application US/10156761
Publication No. US20030119018A1
GENERAL INFORMATION:

APPLICANT: OMURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 12370
LENGTH: 272
TYPE: PRT
ORGANISM: Streptomyces avermitilis
US-10-156-761-12370

Query Match 36.6%; Score 78; DB 15; Length 272;
Best Local Similarity 46.3%; Pred. No. 0.41;
Matches 19; Conservative 7; Mismatches 15; Indels 0; Gaps 0;

QY 1 AKKYAKKAKAEKAKKAYKAAAEKAKYKAEKAAAEKAAKEA 41
DB 90 AAKAKQAKSDLADAKKKAETKKAEAARRAAERAAASRA 130

RESULT 15
US-09-820-843A-24
Sequence 24, Application US/09820843A
Publication No. US20030039963A1
GENERAL INFORMATION:
APPLICANT: Council of Scientific and Industrial Research
TITLE OF INVENTION: A COMPUTATIONAL METHOD FOR THE IDENTIFICATION OF CANDIDATE PROTEI
FILE REFERENCE: Q63915
CURRENT APPLICATION NUMBER: US/09/820,843A
CURRENT FILING DATE: 2001-03-30
NUMBER OF SEQ ID NOS: 118
SOFTWARE: Patencin version 3.0
SEQ ID NO 24
LENGTH: 309
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: polyh/droxylalkanoate synthesis protein Phaf
NAME/KEY: misc feature
OTHER INFORMATION: g1|9951352
US-09-820-843A-24

Query Match 36.6%; Score 78; DB 11; Length 309;
Best Local Similarity 54.8%; Pred. No. 0.47;
Matches 23; Conservative 0; Mismatches 19; Indels 0; Gaps 0;

QY 1 AKKYAKKAKAEKAKKAYKAAAEKAKYKAEKAAAEKAAKEA 42
DB 142 AKAAAKPAKPAKPAKPAKTAAPAAKPAKAAKPAKPAKPA 183

Search completed: January 21, 2004, 09:10:08
Job time : 16.3636 secs

US-09-405-743A-7

[illegible]

Sequence 5094, Application US/09107532A
Patent No. 6583275
GENERAL INFORMATION:
APPLICANT: Lynn A Doucette-Stamm and David Bush
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 7310
CORRESPONDENCE ADDRESS:
ADDRESSER: GENOME THERAPEUTICS CORPORATION
STREET: 100 Beaver Street
CITY: Waltham
STATE: Massachusetts
COUNTRY: USA
ZIP: 02354
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: PC
OPERATING SYSTEM: <Unknown>
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107,532A
FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/085,598
FILING DATE: 14 May 1998
APPLICATION NUMBER: 60/051571
FILING DATE: JULY 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Arinello, Pamela Deneke
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 5094:
SEQUENCE CHARACTERISTICS:
LENGTH: 700 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium
FEATURE:
NAME/KEY: misc_feature
LOCATION: (B) LOCATION 1...700
SEQUENCE DESCRIPTION: SEQ ID NO: 5094:
US-09-107-532A-5094

Query Match 36.4%; Score 77.5; DB 4; Length 700;
Best Local Similarity 53.7%; Pred. No. 0.55;
Matches 22; Conservative 7; Mismatches 9; Indels 3; Gaps 2;

QY 6 KKAKEKAKKAYKAAKAKKAYE-KAAEKAAKAAVEA 45
DB 542 KEKKEEEKK-KAEAKKAAEOKAQAQAAAAQAQA 580

RESULT 15
US-09-041-889-39
Sequence 39, Application US/09041889
Patent No. 6033864
GENERAL INFORMATION:
APPLICANT: Braun, Jonathan
APPLICANT: Cohavy, Ofer
TITLE OF INVENTION: Diagnosis, Prevention and Treatment of
TITLE OF INVENTION: Ulcerative Colitis, and Clinical Subtypes Thereof, Using
TITLE OF INVENTION: Microbial UC PANCA antigens
NUMBER OF SEQUENCES: 41
CORRESPONDENCE ADDRESS:
ADDRESSER: Campbell & Flores LLP
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego

STATE: California
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/041,889
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/837,058
FILING DATE: 11-APR-1997
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-PM 3006
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 39:
SEQUENCE CHARACTERISTICS:
LENGTH: 103 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-041-889-39

Query Match 36.2%; Score 77; DB 3; Length 103;
Best Local Similarity 52.2%; Pred. No. 0.088;
Matches 24; Conservative 1; Mismatches 13; Indels 8; Gaps 1;

QY 1 AKKYAKKAKAKA-----KKAAYKAAEAKKAKYKAAAEKAA 38
DB 51 AKPKAKKAGAKAKKPPGATPKKAKKAAAGAKKAVKTPKAKKPPA 96

Search completed: January 21, 2004, 09:02:20
Job time : 8.47403 secs

A:Reference number: A25550; MUID:87040778; PMID:3022245
A:Accession: A25550
A:Molecule type: DNA
A:Residues: 1-210 <KNO>
A:Cross-references: GB:J04488; MID:g9616; PIDN:CA28177.1; PID:g9617
C:Superfamily: histone H1
C:Keywords: chromosomal protein; DNA binding; nucleosome; nucleu

Query Match 34.9%; Score 181; DB 2; Length 210;
Best Local Similarity 50.0%; Pred. No. 3.2e-05;
Matches 58; Conservative 8; Mismatches 36; Indels 14; Gaps 4;

Qy 2 KCTAKKAEKAYAKKAEKKAYAKKAYAAE---AKKAKAEKKAYAKKAEAKK 57
Db 96 KTEKAKA-RAAKKAKAKAAKKKEKKCAATKARKETLAKKAKKAKKAYKKPAKAK 154
58 KEAKKAAKKYAKKAAKEKKEYYAAAEAKKA-----EAAKAYKAAKAAKAAKAA 106
155 KPAAKAKKAPKAAK--KAAKKPAAKKPAAKKAKKAPKAAKPAKAAKPAKAAKAA 208

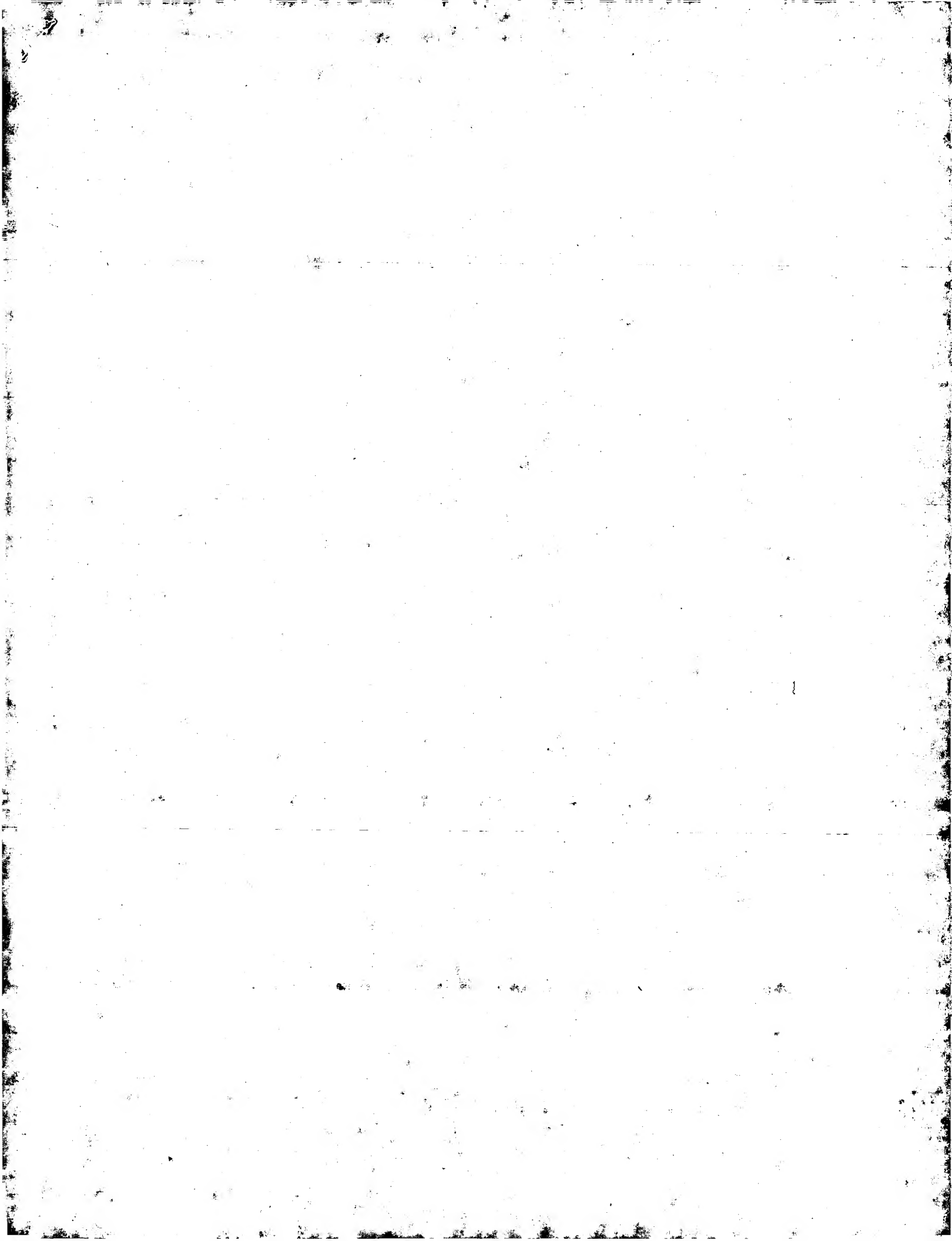
RESULT 5
E83525
Tola protein PA0971 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: E83525
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; B
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kae, A.; Larbig, K.; Lam
a, Lory, S.; Olson, M.V.
Nucleotide 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic path
A:Reference number: A82950; MUID:20437337; PMID:10984043
A:Accession: E83525
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-347 <STO>
A:Cross-references: GB:AE004530; GB:AE004091; MID:g9946865; PIDN:AAG04360.1; GSPDB:GN00
A:Experimental source: strain PA01
C:Genetics:
A:Gene: tola, PA0971

Query Match 34.5%; Score 179; DB 2; Length 347;
Best Local Similarity 49.1%; Pred. No. 6.2e-05;
Matches 54; Conservative 17; Mismatches 29; Indels 10; Gaps 4;

Qy 6 KKAERAYAKKAKAKAKK-----KAYAKKEKAYAAAKKKKAKAEKAYAKKAEAKK 59
Db 99 QKLEQOQVAAAKAEOKKADARAKAEQKAAEAKKADAKAAEAKAEKKQADIAKRR 158
60 AYKAEAKKAYAKKAEKKEVAAAEKKAKEAKKAYAAKAAKAAKAAKAAEAAVEA 109
159 A-EDEAKK--KAAEDAKKK-AAEDAKKAAEAKKAAEAKKAAEAKKAAEAAVEA 204

RESULT 6
AG0592
C:Species: Salmonella enterica subsp. enterica serovar Typhi (strain CT18)
C:Note: This species has also been called Salmonella typhi
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C:Accession: AG0592
R:Parikh, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher
h, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar
S.; Moule, S.; O'Garra, P.
Nucleotide 413, 848-852, 2001
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K
A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica sero
A:Reference number: AB0502; MUID:21534947; PMID:11677608
A:Accession: AG0592
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-576 <PAR>

Search completed: January 21, 2004, 09:01:40
Job time : 20.8182 secs




```

RT  L10ubae R.;;
RA  "TolA central domain interacts with Escherichia coli porins.";
RL  EMOO J. 15:6408-6415(1996).
RM  (6)
RP  X-RAY CRYSTALLOGRAPHY (1.85 ANGSTROMS) OF 298-421.
RX  MEDLINE=99332679; PubMed=10404600;
RA  Lubkowski J., Hennecke F., Plueckhuhn A., Wlodawer A.;
RT  "Flamentous phase infection: crystal structure of g3p in complex
  with its coreceptor, the C-terminal domain of TolA.";
RL  Structure 7:711-722(1999).
CC  -1- FUNCTION: INVOLVED IN THE TONB-DEPENDENT UPTAKE OF GROUP A
CC  COLICINS (COLICINS A, E1, E2, E3, AND K). NECESSARY FOR THE
CC  COLICINS TO REACH THEIR RESPECTIVE TARGETS AFTER INITIAL
CC  BINDING TO THE BACTERIA. ALSO INVOLVED IN THE TRANSLLOCATION
CC  OF BACTERIOPHAGE DNA.
CC  -1- SUBUNIT: INTERACTS, VIA DOMAIN II, WITH PORINS OMPc, OMPc, PHOE
CC  AND LAMB.
CC  -1- SUBCELLULAR LOCATION: Type II membrane protein. Inner membrane.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; M28232; AAA24683.1; -.
DR  EMBL; AE000177; AAC73833.1; -.
DR  EMBL; D90713; BAA35405.1; -.
DR  PIR; JY0057; JY0057.
DR  PDB; 1TOL; 20-MAY-99.
DR  Ecogene; EG11007; tolA.
KW  Transport; Protein transport; Bacteriocin transport; Transmembrane;
KW  Repeat; Inner membrane; 3D-structure; Complete proteome.
FT  DOMAIN 1 13 CYTOPLASMIC (POTENTIAL).
FT  TRANSMEM 14 34 POTENTIAL.
FT  DOMAIN 35 421 PERIPLASMIC (POTENTIAL).
FT  DOMAIN 48 310- DOMAIN II (ALPHA-HELICAL).
FT  DOMAIN 311 421 DOMAIN III (FUNCTIONAL).
FT  DOMAIN 224 278 10 X TANDEM REPEATS OF [ED]-K(1,2)-
FT  A(2,4).
FT  HELIX 335 349
FT  TURN 350 351
FT  TURN 353 354
FT  HELIX 355 358
FT  TURN 359 360
FT  STRAND 363 369
FT  TURN 371 372
FT  STRAND 375 383
FT  HELIX 385 397
FT  HELIX 406 412
FT  TURN 413 414
FT  STRAND 416 421
SQ  SEQUENCE 421 AA; 43156 MW; 8B2F52B4B97C655E CRC64;
Query Match 36.5%; Score 189.5; DB 1; Length 421;
Beat Local Similarity 53.0%; Pred. No. 4.6e-06;
Matches 61; Conservative 11; Mismatches 32; Indels 11; Gaps 5
QY 1 AKKAKKAE---KAYAKKAKAAKKKAYAKKEAAVYKAALEKKKAKAEAKKYAK-EAAK 55
DB 120 AEEAAKQAEELKQKQAEAAAKAAAKAAAKAAEADAKA--AEEAAKKAADDAKKKAAEAAEK 177
QY 56 AKKEAY-AEAAKVKYAKAAAEKKEEVAABAKK---AAEAAKVKAAAKYAAAKAEAA 106
DB 178 AAEEKQKAEAAALAKKKAEEAAEAAEAAEAAEAAKKAATTEAAEAKKAAEKKAAEKA 232

```

```

DR 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 15-JUN-1999 (Rel. 38, Last annotation update)
DS Late histone H1.
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC Echinoidea; Euechinoidea; Echinacea; Temnopneustoidae;
OC Lytechinus.
XX NCBI_TaxID=7653;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RX MEDLINE=87040778; PubMed=3022245;
RA Knowles J.A., Childs G.J.;
RT "Comparison of the late H1 histone genes of the sea urchins
RL Lytechinus pictus and Strongylocentrotus purpuratus.",
RT Nucleic Acids Res. 14:8121-8133(1986).
CC -1- FUNCTION: HISTONES H1 ARE NECESSARY FOR THE CONDENSATION OF
CC NUCLEOSOME CHAINS INTO HIGHER ORDER STRUCTURES.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: BELONGS TO THE HISTONE H1/H5 FAMILY.
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its use by
CC CC non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (see http://www.1db-sdb.ch/announce/
CC CC or send an email to license@1db-sdb.ch).
CC -----
DR EMBL: X04488; CAA28177.1; -.
DR PIR: A25550; A25550.
DR HSSP: P02259; 1HST.
DR InterPro: IPR005818; Histone_H1/H5.
DR InterPro: IPR005819; Histone_H5.
DR InterPro: IPR003216; Linkerhist_N.
DR Pfam: PF00538; linker histone; I.
DR PRINTS: PR00634; HISTONH5.
DR Prodom: PD000373; Linkerhist_N; 1.
DR SMART: SM00526; H1S; 1.
DR K01 Chromosomal protein, Nuclear protein; DNA-binding; Multigene family.
KW SEQUENCE 210 AA; 21746 MW; 08C38F64894007E2 CRC64;
SQ
Query Match 34.9%; Score 181; DB 1; Length 210;
Best Local Similarity 50.0%; Pred. No. 9.3e-06;
Matches 58; Conservative 8; Mismatches 36; Indels 14; Gaps 4
QY 2 KKVAKKEKAAVAKKAAKEKKAAVAKKAAVKA-----AKKAKAAEAKKYAKAAK 57
DB 96 KTEAOKA-RAAAKKAATLAAKKGEKKAAATTKARKEKTLAAKKAAKKAAKKYKKAATAK 154
QY 58 KEAVYAAEKKYAKAKAKAEKKEVAAAIAAKKA-----EAAKAYAAEAAKAAAKAA 106
DB 155 KPAKKAAKKPAK--KAKKPAKKPAKKPAKKPAKKPAKKAAKPAKKAAKPAKKAAK 208
RESULT 3
TOLA_PSEAE
AC P50600; STANDARD; PRT; 347 AA.
DT 01-OCT-1996 (Rel. 34, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE TOLA protein.
GN TOLA OR PA0971.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PAO;
RX MEDLINE=97113525; PubMed=8955385;

```

RA Dennis J.J., Lafontaine E.R., Sokol P.A.;
 RT "Identification and characterization of the tolQRA genes of
 RL Pseudomonas aeruginosa.";
 RA J. Bacteriol. 178:7059-7068 (1996).
 (2)
 RP REVISIONS TO N-TERMINUS.
 RA Duan K., Sokol P.A.;
 RL Submitted (AUG-1999) to the EMBL/Genbank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 15692 / PAOI;
 RX MEDLINE=20437337; PubMed=10984043;
 RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
 RA Hickey M.J., Brinkman F.S.L., Hutnagle W.O., Kowalik D.J., Lagrou M.,
 RA Gader R.L., Goltz L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
 RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Laidig K., Lam R.M.,
 RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
 RA Reizer J., Sater M.H., Hancock R.E.W., Lory S., Olson M.V.;
 RT "Complete genome sequence of Pseudomonas aeruginosa PAOI, an
 RT opportunistic pathogen.";
 RL Nature 406:959-964 (2000).
 CC -1- FUNCTION: INVOLVED IN THE TONB-INDEPENDENT UPTAKE OF PROTEINS
 CC (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Inner membrane
 CC (Potential).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U39558; AAC44660.2; -.
 DR EMBL: AB004530; AAG04360.1; -.
 DR PIR: E83525; E83525.
 DR InterPro: IPR006260; TonB_C.
 DR TIGFAPs: TIG01352; tonB_Cterm; 1.
 KW Transport; Protein transport; Transmembrane; Repeat; Inner membrane;
 KW Complete proteome.
 FT DOMAIN 1 16 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 17 37 POTENTIAL.
 FT DOMAIN 38 347 PERIPLASMIC (POTENTIAL).
 FT DOMAIN 209 216 POLY-ALA.
 SQ SEQUENCE 347 AA; 37935 MW; EBD4B04AA095945 CRC64;
 Query Match 34.5%; Score 179; DB 1; Length 347;
 Best Local Similarity 49.1%; Pred. No. 1.9e-05;
 Matches 54; Conservative 17; Mismatches 23; Indels 10; Gaps 4;
 Oy 6 KKAQKAVAKKAAKAKK-----KAYAKKAAKAVAAKAKKAAKAKKAAKAKKAAKAKK 59
 Db 99 QKLEQQQVAAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAK 158
 Oy 60 AYKAAKAYKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAK 109
 Db 159 A-EDEAKK--KAAEDAKKK-AAEDAKKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAK 204
 RESULT 4
 MST1_DROHY STANDARD; PRT; 344 AA.
 ID MST1_DROHY
 AC Q08695;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Axoneme-associated protein met101(1).
 GN MST101(1).
 OS Drosophila hydei (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydriidae; Drosophilidae; Drosophila.

OX NCBI_TaxID=7224;
 RN [1]
 RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
 RC TISSUE=Testis;
 RX MEDLINE=94200512; PubMed=8150205;
 RA Neesen J., Buemann H., Heinlein U.A.;
 RT "The Drosophila hydei gene Dhmst101(1) encodes a testis-specific,
 RT repetitive, axoneme-associated protein with differential abundance in
 RT Y chromosomal deletion mutant flies.";
 RL Dev. Biol. 162:414-425 (1994).
 CC -1- FUNCTION: POSSIBLE STRUCTURAL ROLE IN THE SPERM TAIL. IT IS
 CC ASSOCIATED WITH AXONEMAL STRUCTURES.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -1- TISSUE SPECIFICITY: TESTIS. LOCATED IN SPERMATOCTES AND
 CC SPERMATID BUNDLES.
 CC -1- DOMAIN: THE PREDOMINANT STRUCTURE IS ALPHA-HELICAL.
 CC -1- POLYMORPHISM: THE NUMBER OF REPEATS VARIES BETWEEN STRAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X73480; CAAS1875.1; -.
 DR PIR: S34153; S34153.
 DR FlyBase: FBgn011816; Dhyd\met101(1).
 KW Sperm; Repeat; Multigene family.
 FT DOMAIN 58 337
 FT SEQUENCE 344 AA; 37793 MW; 24C65D2510387E2A CRC64;
 Query Match 33.8%; Score 175.5; DB 1; Length 344;
 Best Local Similarity 49.2%; Pred. No. 3.1e-05;
 Matches 58; Conservative 8; Mismatches 35; Indels 17; Gaps 5;
 Oy 1 AKKTAQKAEK---AYAKKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAK 49
 Db 65 AKKEKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAK 124
 Oy 50 AKKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAK 106
 Db 125 CAEAAKKEKA--AEKKCAEAAKKEK--AAEKCAEAAKKEKAAKAAKCAEAA 177
 RESULT 5
 MST2_DROHY STANDARD; PRT; 1391 AA.
 ID MST2_DROHY
 AC Q08696;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Axoneme-associated protein met101(2).
 GN MST101(2).
 OS Drosophila hydei (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydriidae; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7224;
 RN [1]
 RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
 RX MEDLINE=9504538; PubMed=7957199;
 RA Neesen J., Padmanabhan S., Buemann H.;
 RT "Randomly arranged repeats of a novel highly charged 16-amino-acid
 RT motif representing the major component of the sperm-tail-specific
 RT axoneme-associated protein family Dhmst101 form extended
 RT alpha-helical rods within the extremely elongated spermatozoa of
 RT Drosophila hydei.";
 RL Eur. J. Biochem. 225:1089-1095 (1994).
 CC -1- FUNCTION: POSSIBLE STRUCTURAL ROLE IN THE SPERM TAIL.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.

```
CC -1- TISSUE SPECIFICITY: TESTIS. PRIMARY SPERMATOCYTES AND EARLY
CC SPERMATIDS.
CC -1- DOMAIN: THE PREDOMINANT STRUCTURE IS ALPHA-HELICAL.
CC -1- POLYMORPHISM: LENGTH POLYMORPHISMS EXIST BETWEEN DIFFERENT
CC STRAINS, MOST LIKELY CAUSED BY LENGTH VARIATIONS WITHIN THE TANDEM
CC REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb.ch/announce/
CC or send an email to license@isb.stb.ch).
CC -----
DR EMBL; X73481; CAA51876.1; -.
DR PIR; S51364; S51364.
DR FLYBase; FBgn0020733; Dhyd\mst101(2).
KW Sperm; Repeat; Multigene family; Polymorphism.
FT FT 332 1268 59 x 16 AA APPROXIMATE TANDEM REPEATS OF
FT DOMAIN [KR]-K-X-C-X-X-X-A-K-X-K-X-X-E.
SQ SEQUENCE 1391 AA; 159000 MW; 1B2A368F30F48878 CRC64;
Query Match 33.0%; Score 171.5; DB 1; Length 1391;
Best Local Similarity 46.2%; Pred. No. 0.00017;
Matches 54; Conservative 11; Mismatches 37; Indels 15; Gaps 4;
Qy 1 AKKYAKKAEEKAY---AKKATAAEKKAYAKEKAYAYAAE-----AKKKAAVEARKY 49
Db 1068 AKKLKEAEKKCKCEBRACKKEAEKCKCEBRACKLKLEAEKCKCEBRACKKEAEKRR 11227
Qy 50 AKEAAKAKKENYAKELKYAAQAARKEKEYAAAEEKKRAEKBAKAYAAEAANKAAAEAA 106
Db 1128 CEBAKREKKA--AEKKCAEAARKEKE--ATEKQCAEAARKEKAEKRRCAEAA 1180
RESULT 6
H1_PARAN STANDARD; PRT; 248 AA.
AC P03256;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Histone H1, gonadal.
OS Parachinus angulosus (Angulate sea urchin).
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC Echinoidae; Euechinoidae; Echinacea; Echinoida; Echinidae;
OC Parachinus.
CX NCBI_TaxID=7658;
RN [1]
RP SEQUENCE OF 1-84.
RX MEDLINE=80156831; PubMed=6767609;
RA Strickland W.N., Strickland M., de Groot P.C., von Holt C.,
RA Wittmann-Liebold B.;
RT "The primary structure of histone H1 from sperm of the sea urchin
RT Parachinus angulosus. 1. Chemical and enzymatic fragmentation of the
RT protein and the sequence of amino acids in the four N-terminal
RT cyanogen bromide peptides."
RL Eur. J. Biochem. 104:559-566(1980).
RN [2]
RP SEQUENCE OF 80-248.
RX MEDLINE=80156832; PubMed=7367905;
RA Strickland W.N., Strickland M., Brandt W.F., von Holt C., Lehmann A.,
RA Wittmann-Liebold B.;
RT "The primary structure of histone H1 from sperm of the sea urchin
RT Parachinus angulosus. 2. Sequence of the C-terminal CNBr peptide and
RT the entire primary structure."
RL Eur. J. Biochem. 104:567-578(1980).
CC -1- FUNCTION: HISTONES H1 ARE NECESSARY FOR THE CONDENSATION OF
CC NUCLEOSOME CHAINS INTO HIGHER ORDER STRUCTURES.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- TISSUE SPECIFICITY: SPERM.
CC -1- SIMILARITY: BELONGS TO THE HISTONE H1/H5 FAMILY.
```

```

DR PR1; A91090; HSTRIP.
DR HSSP; P02259; HIST.
DR InterPro; IPR005818; Histone_H1/H5.
DR InterPro; IPR005819; Histone_H5.
DR InterPro; IPR003216; Linkerhist_N.
DR Pfam; PF00538; linker_histone_1.
DR PRINTS; PR00624; HISTONEH5.
DR ProDom; PD000373; Linkerhist_N; 1.
DR SMART; SMO0526; H15; 1.
KW Chromosomal protein; Nuclear protein; DNA-binding; Multigene family;
KW Sperm.
FT VARIAT 144 144 K -> R.
SQ SEQUENCE 248 AA; 26387 MW; 1B25B3F136541947 CRC64;

Query Match 32.9%; Score 170.5; DB 1; Length 248;
Best Local Similarity 46.0%; Pred. No. 5e-05;
Matches 52; Conservative 9; Mismatches 41; Indels 11; Gaps 2;

QY 2 KTVAKKEKVVAKYAKKAKEKKYAKKEAAVKAAEAKKCAKEAKYAKAEAKKAEAY 61
Db 120 KKAKGTSAAKKAKKAKKAAKAKKAKKAKKAAKKAALKKKAAAKRDAAKKAKKAPKK 179

QY 62 KA--EARKAK-----AAKKEKYAAAEKKKAADAKAYAEADAKAAK 103
Db 180 KAAKAKKAPKAPKSPKAKKPKAKKPKKAKRSPKAKKAKAAGKRRKPAKAKAR 232

RESULT 7
H1B_STRPU STANDARD; PRT; 211 AA.
ID H1B_STRPU
AC P15B69;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Histone H1-beta, late embryonic.
OS Strongylocentrotus purpuratus (Purple sea urchin).
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinzoa;
OC Echinoidae; Euechinoida; Echinacea; Echinoida; Strongylocentrotidae;
OC Strongylocentrotus.
OX NCBI_TaxID=7668;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88246461; PubMed=2837660;
RA Lai Z.-C., Childs G.;
RT "Characterization of the structure and transcriptional patterns of
RT the gene encoding the late histone subtype H1-beta of the sea urchin
RT Strongylocentrotus purpuratus.";
RL Mol. Cell. Biol. 8:1842-1844(1988).
CC -1- FUNCTION: HISTONES H1 ARE NECESSARY FOR THE CONDENSATION OF
CC NUCLEOSOME CHAINS INTO HIGHER ORDER STRUCTURES.
CC NUCLEOSOME LOCATION: Nuclear.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: BELONGS TO THE HISTONE H1/H5 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/1sb-sib.ch).
CC CC send an email to license@1sb-sib.ch.
CC -----
CC EMBL; M20314; AAA30052.1; -.
DR DR PIR; A28100; A28100.
DR HSSP; P02259; HIST.
DR InterPro; IPR005818; Histone_H1/H5.
DR InterPro; IPR005819; Histone_H5.
DR InterPro; IPR003216; Linkerhist_N.
DR Pfam; PF00538; linker_histone_1.
DR PRINTS; PR00624; HISTONEH5.
DR ProDom; PD000373; Linkerhist_N; 1.
DR SMART; SMO0526; H15; 1.
KW Chromosomal protein; Nuclear protein; DNA-binding; Multigene family.
SQ SEQUENCE 211 AA; 22169 MW; 9F214581334ABE7A CRC64;

```



```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X02624; CAB37646.1; -.
CC PIR: A02584; HSTRIR.
CC HSSP: P08287; IGHC.
CC InterPro: IPR005818; Histone_H1/H5.
CC InterPro: IPR005819; Histone_H5.
CC Pfam: PF00538; linker histone; 1.
CC PRINTS: PR00624; HISTONEH5.
CC SMART: SM00526; H15; 1.
CC Chromosomal protein; Nuclear protein; DNA-binding; Multigene family;
CC Acetylation.
CC INIT MET 0 0 ACETYLATION (BY SIMILARITY).
CC MOD RES 1 1 GLOBULAR.
CC DOMAIN 27 100
CC SEQUENCE 206 AA; 20672 MW; 72C440798066716C CRC64;
CC
CC Query Match 30.9%; Score 160.5; DB 1; Length 206;
CC Best Local Similarity 47.7%; Pred. No. 0.00019;
CC Matches 52; Conservative 9; Mismatches 35; Indels 13; Gaps 4;
CC
QY 1 AKKVAK-----KAEKAYAKAKAKAKKAYAKKAEKAYAKAEKAKKAKAKKAYAKAEKAK 55
DB 105 AKKPAKAAAPKAKKAAKPAKAAKPAKAAKPAKAAKPAKAAKSPKAKKAPTPKAAKSPK 164
QY 56 AKKEAYKAEKAYAKAKAKAEKAYAAAEKAEKAEKAYAKAEKAAKAEKAAKAE 104
DB 165 VKKPA--AAAK---KAAKSPK--ATKAAKPAKAAKPAKAAKAAKSPK 205

```

```

RA Wild A.;
RL Submitted (DEC1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: HISTONES H1 ARE NECESSARY FOR THE CONDENSATION OF
CC NUCLEOSOME CHAINS INTO HIGHER ORDER STRUCTURES.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: BELONGS TO THE HISTONE H1/H5 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X83509; CAA58498.1; -.
CC EMBL: AF531304; AAN06704.1; -.
CC PIR: Z98744; CAB11421.1; -.
CC PIR: S51660; S51660.
CC HSSP: P08287; IGHC.
CC Genew; HGNC:4719; HISTH1B.
CC MIM: 142711; -.
CC GO: GO:0005718; C:nucleosome; NAS.
CC GO: GO:0003677; F:DNA binding activity; NAS.
CC GO: GO:0007001; P:chromosome organization and biogenesis (see. . .; NAS.
CC GO: GO:0006334; P:nucleosome assembly; NAS.
CC DR InterPro: IPR005818; Histone_H1/H5.
CC DR InterPro: IPR005819; Histone_H1/H5.
CC Pfam: PF00538; linker histone; 1.
CC PRINTS: PR00624; HISTONEH5.
CC SMART: SM00526; H15; 1.
CC Chromosomal protein; Nuclear protein; DNA-binding; Multigene family;
CC Acetylation.
CC INIT MET 0 0
CC MOD RES 1 1 ACETYLATION.
CC CONFLICT 215 217 MISSING (IN REF. 1).
CC SEQUENCE 225 AA; 22449 MW; 26CD4A1B5D463CDA CRC64;
CC
CC Query Match 30.9%; Score 160.5; DB 1; Length 225;
CC Best Local Similarity 46.8%; Pred. No. 0.0002;
CC Matches 51; Conservative 7; Mismatches 44; Indels 7; Gaps 2;
CC
QY 1 AKKVAKAEKAYAKAKAKAEKAYAKAEKAYAKAEKAEKAYAKAEKAEKAEKAEKAEKAEKAE 55
DB 118 AKPKAKKGAAPKAKKPAKATPKK--AKAAGAKKAVKTPKPAKPAAGVKKAAKSPK 175
QY 56 AKKEAYKAEKAYAKAKAEKAYAAAEKAEKAEKAYAKAEKAAKAEKAAKAE 104
DB 176 AKAAKPKKATKSPKPAKPAKPAKPAKPAKPAKPAKPAKPAKAAK 224

```

```

RESULT 14
DBH_MYCSM STANDARD; PRT; 208 AA.
AC Q9ZEC5;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DNA-binding protein HU homolog (Histone-like protein) (H1p).
GN HUP OR HLP
OS Mycobacterium smegmatis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1772;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 700084 / mc(2)155;
RA MEDLINE=9110209; PubMed=9894918;
RA Lee B.H., Murugaou-Oei B., Dick T.;
RT "Upregulation of a histone-like protein in dormant Mycobacterium
RT smegmatis.";
RL Mol. Gen. Genet. 260:475-479(1998).
CC -!- FUNCTION: THIS PROTEIN BELONGS TO THE HISTONE LIKE FAMILY OF

```

```

CC PROKARYOTIC DNA-BINDING PROTEINS WHICH ARE CAPABLE OF WRAPPING
CC DNA TO STABILIZE IT AND PREVENT ITS DENATURATION UNDER EXTREME
CC ENVIRONMENTAL CONDITIONS (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE BACTERIAL HISTONE-LIKE PROTEIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF068138; AAD13809.1; -.
DR HSSP; P02346; IHUO.
DR InterPro; IPR000119; Bac DNAbind.
DR Pfam; PF00216; Bac DNA binding; 1.
DR ProDom; PD000945; Bac DNAbind; 1.
DR SMART; SM00411; BHL; 1.
DR PROSITE; PS00045; HISTONE-LIKE; 1.
DR DNA-binding; DNA condensation; Repeat.
FT DOMAIN 1 90 BACTERIAL HISTONE-LIKE DOMAIN.
FT DOMAIN 101 205 DEGENERATE REPEATS REGION.
SQ SEQUENCE 208 AA; 21230 MW; CASP577F61F7EP09 CRC64;

Query Match 30.8%; Score 160; DB 1; Length 208;
Best Local Similarity 54.3%; Pred. No. 0.00021;
Matches 57; Conservative 3; Mismatches 31; Indels 14; Gaps 5;

QY 1 AKYVAKKAEEKYAKKAAEKKAAYKAAEAKEKKAAYKAAEAKEKKAAYKAAEAKEKKA 60
DB 111 AKGAAKKAP--AKGAAKKTATKAAAKAPAKKAA-----TKAPAKKATK-APAKKA 161
QY 61 YKAAKKYAKAQAEEKEVAAAEEAKKAQAAYKAAEAQAQAEEA 105
DB 162 TKAPAKKATKAPA-KKAAAKAPAKKA-----ATKAPAKKAAKKA 201

RESULT 15
H1R CHICK STANDARD; PRT; 218 AA.
ID H1R CHICK P08288;
AC 01-AUG-1988 (rel. 08, Created)
DT 01-AUG-1988 (rel. 08, Last sequence update)
DT 15-SEP-2003 (rel. 42, Last annotation update)
DS Histone H1.11R.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87250632; PubMed=3597432;
RA Coles L.S.; Robins A.J.; Madley L.K.; Wells J.R.E.;
RT Characterization of the chicken histone H1 gene complement.
RT Generation of a complete set of vertebrate H1 protein sequences.;
RL J. Biol. Chem. 262:9656-9663(1987).
CC -1- FUNCTION: HISTONES H1 ARE NECESSARY FOR THE CONDENSATION OF
CC NUCLEOSOME CHAINS INTO HIGHER ORDER STRUCTURES.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: BELONGS TO THE HISTONE H1/H5 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M17020; AAA48790.1; -.
DR PIR; C28456; C28456.

```

```

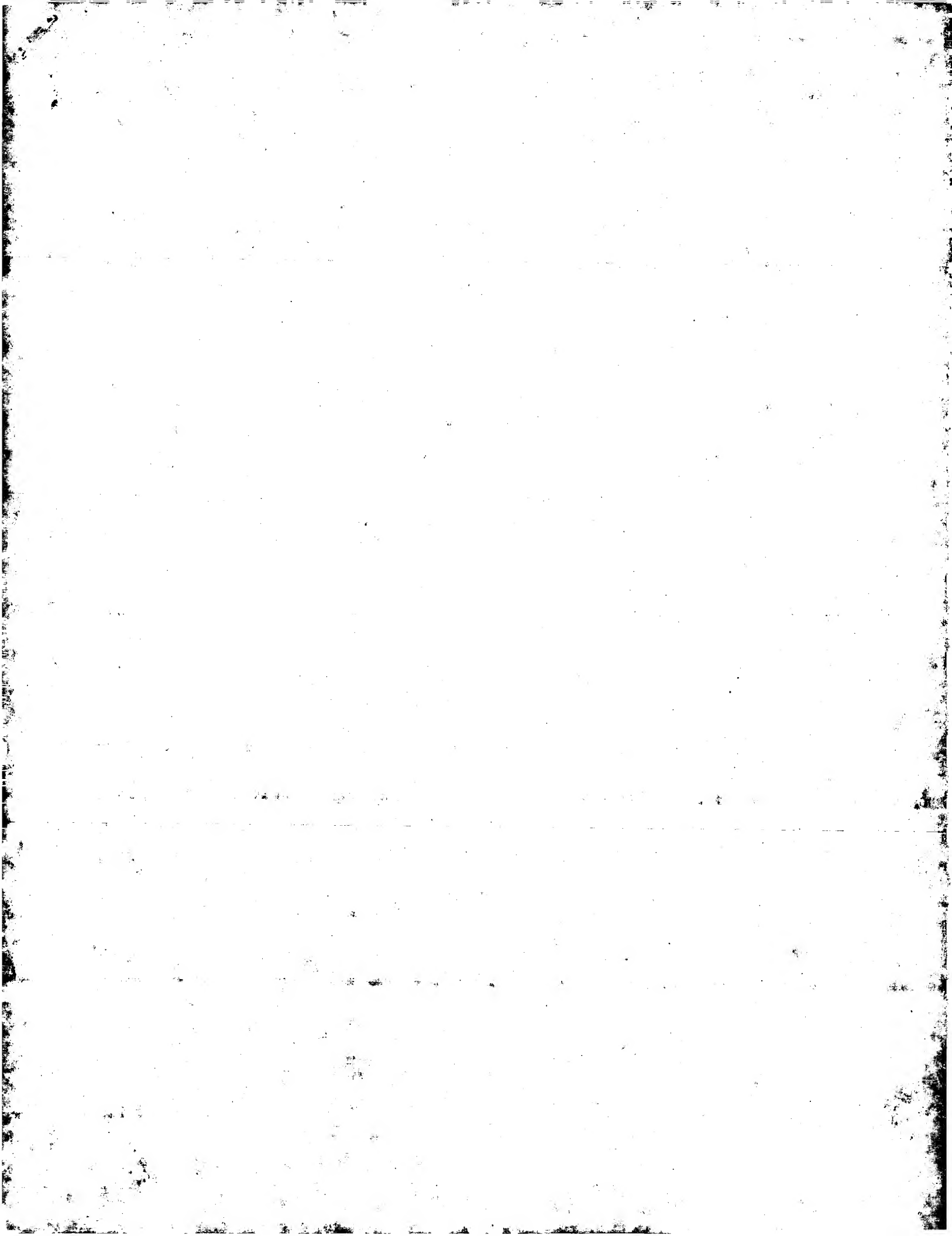
DR PDB; 1GHC; 31-AUG-94.
DR InterPro; IPR005818; Histone_H1/H5.
DR InterPro; IPR005819; Histone_H5.
DR Pfam; PF00538; linker histone; 1.
DR PRINTS; PR00624; HISTONEH5.
DR SMART; SM00526; H15; 1.
DR Chromosomal protein; Nuclear protein; DNA-binding; Multigene family;
KW 3D-structure.
FT INIT_MET 0 0 GLOBULAR.
FT DOMAIN 37 110
SQ SEQUENCE 218 AA; 21672 MW; CB9724BPF14654A6 CRC64;

Query Match 30.7%; Score 159.5; DB 1; Length 218;
Best Local Similarity 47.7%; Pred. No. 0.00023;
Matches 52; Conservative 11; Mismatches 41; Indels 5; Gaps 3;

QY 1 AKYVAKKAEEKYAKKAAEKKAAYAKK---EAYVAKAEEKKAAEKKAEEKYAKKA 57
DB 109 SKKPEGEGEKAPKKKSAKPKKAAAKKPAKAAKKPKKAAVAVKSPKKAKKPAASATKS 168
QY 58 KEAYVAKK-KYAVQAEEKEVAA-AEAKTAEEAKKAYKAAQAQAEE 104
DB 169 VKSPKKAAPKPAVAAKSPAKAKVKKPAKPKAAKPKPAKAKKAAK 217

Search completed: January 21, 2004, 09:00:59
Job time : 15.8636 secs

```



RESULT 2	
ID	Q9WMX1 PRELIMINARY; PRT: 372 AA.
AC	Q9WMX1. 01-NOV-1999 (TREMBLrel. 12, Created)
DT	01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT	01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE	TolA protein.
GN	TOL.
OS	Pseudomonas putida.
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC	Pseudomonadaceae; Pseudomonas.
OX	NCBI_TaxID=303;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=mt-2;
EX	MEDLINE=96198174; PubMed=8626299;
RA	Rodriguez-Herva J.J.; Ramos-Gonzalez M.I.; Ramos J.;
RT	"The Pseudomonas putida peptidoglycan-associated outer membrane
RT	lipoprotein (PAL) is involved in maintenance of the integrity of the
RT	cell envelope."
RL	J. Bacteriol. 178:1699-1706(1996).
RN	[2]
RP	SEQUENCE FROM N.A.
RC	STRAIN=mt-2;
RA	Ramos-Gonzalez J.;
RL	Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
RN	[3]
RP	SEQUENCE FROM N.A.
RC	STRAIN=mt-2;
RA	Rodriguez-Herva J.J.;
RL	Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
RN	[4]
RP	SEQUENCE FROM N.A.
RC	STRAIN=mt-2;
RA	MEDLINE=96422022; PubMed=8824639;
RT	Rodriguez-Herva J.J.; Ramos J.;
RT	"Characterization of an OprX null mutant of Pseudomonas putida."
RL	J. Bacteriol. 178:5836-5840(1996).
DR	EMBL; X74218; CAB50780.1; -;
DR	InterPro; IPR005819; Histone_H5.
DR	InterPro; IPR006260; TonB_C.
DR	PRINTS; PR00624; HISTONEH5.
DR	TIGRfams; TIGR01352; tonB_Cterm; 1.
SO	SEQUENCE 372 AA; 40133 MW; 87649785ECC3C0BC CRC64;
Query Match 36.7%; Score 190.5; DB 2; Length 372;	
Best Local Similarity 50.4%; Pred. No. 6, 3e-06;	
Matches 59; Conservative 14; Mismatches 31; Indels 13; Gaps 4	
QY	5 AKKAERAYAKKAKAAK---EKKAYAKKAKYKAAE-----AKKKAERAKKAKAA 54
DB	118 AKKADA-AKAAEAAKAAEAKKAAEAKKADAKKAAEKQADIAKKKADDAKKAEEBA 176
QY	55 K-AKKEAYKAEKKYAYAKAEKKKYAAAEKKAAEAAAYAAEAAKAAKAAEAYBA 109
DB	177 KKAABEAKKKAABDAKKAABEAKKKAABDAKKAABEAKKKAABDA 233
RESULT 3	
Q8TSC8	PRELIMINARY; PRT: 1866 AA.
AC	Q8TSC8. 01-JUN-2002 (TREMBLrel. 21, Created)
DT	01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT	01-OCT-2002 (TREMBLrel. 22, Last annotation update)
DE	MaebI.
GN	MAEBI.
OS	Plasmodium vivax.
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX	NCBI_TaxID=5855;

```

RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=Salvador:
RA      Michon P., Stevens J.R., Kaneo O., Adams J.H.;
RT      "Evolutionary relationships of conserved cysteine-rich motifs in
RL      adhesive molecules of malaria parasites.";
DR      Mol. Biol. Evol. 0:0-0(2002).
SQ      EMBL: AY042083; AAL10508.1; -
        SEQUENCE 1866 AA; 212420 MW; DC692D7CFAE7D93F CRC64;

Query Match          36.7%; Score 190.5; DB 5; Length 1866;
Best Local Similarity 49.6%; Pred. No. 2,7e-05;
Matches 61; Conservative 17; Mismatches 22; Indels 23; Gaps 6;

OY      5 AKKAEKAYAKKAKAAKAE-----KKAYAKKEAKAYYAAEAATKKA-----KAEAKKYA 50
DB      1262 AKKAE--ARRAAEAKKAEARKEAAKKAEEAAKKAEEARKEARKEARKEAAKKA 1319
OY      51 KEAAKAKKEAAKAAKKYAKAKAAKKEKKEVYAAAEAKKAEAAK---AVYAAEAAKAAAEAA 106
DB      1320 EDARKA--EARKAAEAAKKAEEARKEAAKKAEEARKEAAKKAEEARKEAAKKA--EAA 1376
OY      107 YEAA 109
DB      1377 RKA 1379

RESULT 4
O8FUT1  O8FUT1      PRELIMINARY; PRT; 421 AA.
AC      O8FUT1:
DT      01-MAR-2003 (TREMBLrel. 23, Created)
DT      01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT      01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE      ToLA protein.
GN      ToLA OR C0818.
OS      Escherichia coli O6.
OC      Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC      Enterobacteriaceae; Escherichia.
OX      NCBI_TaxID=217992;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=O6:HI / CFT073 / ATCC 700928;
RX      MEDLINE=22388234; PubMed=12471157;
RA      Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA      Raeko D., Buckles E.J., Liou S.-R., Boutin A., Hackett U., Stroud D.,
RA      Mayhew G.P., Domes D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA      Mobley H.L.T., Rosenberg M.S., Blattner F.R.;
RT      "Extensive mosaic structure revealed by the complete genome sequence
RT      of uropathogenic Escherichia coli.";
RL      Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
DR      EMBL: AE016757; AAN79291.1; -.
KW      Complete proteome.
SQ      SEQUENCE 421 AA; 43184 MW; DB296626F056D385 CRC64;

Query Match          36.3%; Score 188.5; DB 16; Length 421;
Best Local Similarity 53.0%; Pred. No. 9.6e-06;
Matches 61; Conservative 11; Mismatches 32; Indels 11; Gaps 5;

OY      1 AKKAYKAAE-----FAVAKKAAAEKKAAYAKKAEAKKAAKKAEEAKKAAKAAK 55
DB      120 ABEAAKQAEELKQKQAEBAAAAGAAADAKKAFAADKA--AEBAKKAAGAAADAKKAEEAAK 177
OY      56 AKKEAY--KAEAKKYAKAKAEKKEVYAAAEAAK---AEAAKAYKAAKAAKAAEAA 106
DB      178 AAABAQKKAEEAAALAKKKKAAEAAAEAAAEAAKKAATAEAAEKAAAEKAAAEKA 232

RESULT 5
O8ZOT6  O8ZOT6      PRELIMINARY; PRT; 407 AA.
AC      O8ZOT6:
DT      01-MAR-2002 (TREMBLrel. 20, Created)

```

```

RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=Salvador:
RA      Michon P., Stevens J.R., Kaneo O., Adams J.H.;
RT      "Evolutionary relationships of conserved cysteine-rich motifs in
RL      adhesive molecules of malaria parasites.";
DR      Mol. Biol. Evol. 0:0-0(2002).
SQ      EMBL: AY042083; AAL10508.1; -
        SEQUENCE 1866 AA; 212420 MW; DC692D7CFAE7D93F CRC64;

Query Match          36.7%; Score 190.5; DB 5; Length 1866;
Best Local Similarity 49.6%; Pred. No. 2,7e-05;
Matches 61; Conservative 17; Mismatches 22; Indels 23; Gaps 6;

OY      5 AKKAEKAYAKKAKAAKAE-----KKAYAKKEAKAYYAAEAATKKA-----KAEAKKYA 50
DB      1262 AKKAE--ARRAAEAKKAEARKEAAKKAEEAAKKAEEARKEARKEARKEAAKKA 1319
OY      51 KEAAKAKKEAAKAAKKYAKAKAAKKEKKEVYAAAEAKKAEAAK---AVYAAEAAKAAAEAA 106
DB      1320 EDARKA--EARKAAEAAKKAEEARKEAAKKAEEARKEAAKKAEEARKEAAKKA--EAA 1376
OY      107 YEAA 109
DB      1377 RKA 1379

RESULT 4
O8FUT1  O8FUT1      PRELIMINARY; PRT; 421 AA.
AC      O8FUT1:
DT      01-MAR-2003 (TREMBLrel. 23, Created)
DT      01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT      01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE      ToLA protein.
GN      ToLA OR C0818.
OS      Escherichia coli O6.
OC      Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC      Enterobacteriaceae; Escherichia.
OX      NCBI_TaxID=217992;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=O6:HI / CFT073 / ATCC 700928;
RX      MEDLINE=22388234; PubMed=12471157;
RA      Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
RA      Raeko D., Buckles E.J., Liou S.-R., Boutin A., Hackett U., Stroud D.,
RA      Mayhew G.P., Domes D.J., Zhou S., Schwartz D.C., Perna N.T.,
RA      Mobley H.L.T., Rosenberg M.S., Blattner F.R.;
RT      "Extensive mosaic structure revealed by the complete genome sequence
RT      of uropathogenic Escherichia coli.";
RL      Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
DR      EMBL: AE016757; AAN79291.1; -.
KW      Complete proteome.
SQ      SEQUENCE 421 AA; 43184 MW; DB296626F056D385 CRC64;

Query Match          36.3%; Score 188.5; DB 16; Length 421;
Best Local Similarity 53.0%; Pred. No. 9.6e-06;
Matches 61; Conservative 11; Mismatches 32; Indels 11; Gaps 5;

OY      1 AKKAYKAAE-----FAVAKKAAAEKKAAYAKKAEAKKAAKKAEEAKKAAKAAK 55
DB      120 ABEAAKQAEELKQKQAEBAAAAGAAADAKKAFAADKA--AEBAKKAAGAAADAKKAEEAAK 177
OY      56 AKKEAY--KAEAKKYAKAKAEKKEVYAAAEAAK---AEAAKAYKAAKAAKAAEAA 106
DB      178 AAABAQKKAEEAAALAKKKKAAEAAEAAAEAAKKAATAEAAEKAAAEKKAAEKAA 232

RESULT 5
O8ZOT6  O8ZOT6      PRELIMINARY; PRT; 407 AA.
AC      O8ZOT6:
DT      01-MAR-2002 (TREMBLrel. 20, Created)

```


DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
 DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
 DE Tol protein, membrane spanning protein.
 GN TOLA OR STM0747.
 OS Salmonella typhimurium.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Salmonella.
 NCBI_TaxID=602;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
 RX MEDLINE=21534948; PubMed=11677609;
 RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
 Courtney L., Porcullik S., Ali J., Dante M., Du F., Hou S., Layman D.,
 Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
 Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
 Waterston R., Wilson R.K.;
 RA "Complete genome sequence of Salmonella enterica serovar Typhimurium
 LT2.";
 RT Nature 413:852-856(2001).
 RL EMBL; AE008730; AAL19691.1; -;
 DR Complete proteome.
 KW SEQUENCE 407 AA; 41865 MW; 253435211602D75 CRC64;
 SQ
 Query Match 35.5%; Score 184.5; DB 16; Length 407;
 Best Local Similarity 55.5%; Pred. No. 1.7e-05;
 Matches 66; Conservative 11; Mismatches 25; Indels 17; Gaps 8;
 QY 1 AKKYAKKAEKAYAKKAKKAYAKKAKAYK-AAEAKKKAKKAKKAYAK-AKAK 58
 DB 138 AKAAADAKKKAEAAKAAADAK--KKAEEAVKAAADAKKKAEAAKAAADAKKKA 195
 QY 59 EAYK--AEAKYAKAKAKKAYAKKAYAKK--AEAKYAKAEAK--AAKAEAYEA 109
 DB 196 EAAKAAAEKAKKAEAAKAEAKK--AAAEAKKADAEAAKAEAKKADAAAKAAEA 249
 RESULT 6
 Q8X965 PRELIMINARY; PRT; 394 AA.
 AC Q8X965;
 DT 01-MAR-2002 (TReMBLrel. 20, Created)
 DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
 DT 01-OCT-2002 (TReMBLrel. 22, Last annotation update)
 DE Membrane spanning protein, required for outer membrane integrity
 DE (Membrane spanning protein TOLA).
 GN TOLA OR Z0907 OR EC50774.
 OS Escherichia coli O157:H7.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 NCBI_TaxID=83334;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
 RX MEDLINE=21074935; PubMed=11206551;
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
 Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 Postel G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
 Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 Welch R.A., Blattner F.R.;
 RA "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
 RT Nature 409:529-533(2001).
 RL [2]
 RN SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / RIMD 0509952;
 RX MEDLINE=21156231; PubMed=11258796;
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
 Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
 Iida T., Takami H., Honda T., Sasaki A., Shikawa S., Shiba T.,
 Kuhara S., Shiba T., Hattori M., Shingawa H.;
 RA "Complete genome sequence of enterohaemorrhagic Escherichia coli
 O157:H7 and genomic comparison with a laboratory strain K-12.";

RL DNA Res. 8:11-22(2001).
 DR EMBL; AE005252; AAG55075.1; -;
 DR EMBL; AF002553; BAB34197.1; -;
 DR InterPro; IPR00104; Antifreeze_1.
 DR PRINTS; PR00308; ANTIFREEZE1.
 KW Complete proteome.
 SQ SEQUENCE 394 AA; 40517 MW; 5858D8E8230BDE28 CRC64;
 Query Match 35.0%; Score 181.5; DB 16; Length 394;
 Best Local Similarity 52.2%; Pred. No. 2.7e-05;
 Matches 60; Conservative 12; Mismatches 32; Indels 11; Gaps 5;
 QY 1 AKKYAKKAEKAYAKKAKKAYAKKAKAYKAAEAKKKAKKAYAK-EAK 55
 DB 120 AEEAKKQELKQKAEAAKAAADAKKAEADKA-AEAAKAAADAKKAEAAK 177
 QY 56 AKKAEY-KAEAKYAKAKAKKAYAKKAYAKK--AEAKYAKAEAAKAAEA 106
 DB 178 AAEEAKKAEAAKAAALKKKAEAAEAAEAAKAAAEKAAADKKAEEKAAEKKA 232
 RESULT 7
 Q828C1 PRELIMINARY; PRT; 376 AA.
 AC Q828C1;
 DT 01-MAR-2002 (TReMBLrel. 20, Created)
 DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
 DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
 DE Tol protein.
 GN STY0793.
 OS Salmonella typhi.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Salmonella.
 NCBI_TaxID=601;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CT18;
 RX MEDLINE=21534947; PubMed=11677608;
 RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
 Baker S., Baaham D., Bentley S.D., Holden M.T.G., Sebaiha M.,
 Churcher C., Mungall K.L., Berridge K., Chillingworth T., Connor P.,
 Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
 Fellwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jorgensen K.,
 Krogan A., Larsen T.S., Leather S., Moule S., O'Garra P., Parry C.,
 Quail M., Rutherford K., Skellern M., Skellern J., Stevens K.,
 RA Whitehead S., Barrett B.G.;
 RA "Complete genome sequence of a multiple drug resistant Salmonella
 RT enterica serovar Typhi CT18.";
 RL Nature 413:848-852(2001).
 DR EMBL; AL627268; CAD05209.1; -;
 KW Complete proteome.
 SQ SEQUENCE 376 AA; 38804 MW; EC21P2C47678A842 CRC64;
 Query Match 34.3%; Score 178; DB 16; Length 376;
 Best Local Similarity 52.3%; Pred. No. 4.5e-05;
 Matches 58; Conservative 10; Mismatches 33; Indels 10; Gaps 4;
 QY 1 AKKYAKKAEKAYAKKAKKAYAKKAKAYKAAEAKKKAKKAYAK-EAK 60
 DB 137 AKAAADAKKKAEAAKAAADAKKAEAAEAA--AKAAADAKKAEAAEAA--AKAAAEAKK-- 190
 QY 61 YKAEAKYAKAKAKKAYAKKAYAKKAEAAKAAEAAKAAEAAKAAEAA--KEAYE 108
 DB 191 -KAEAAKAAADAKKADAEAAKAAEAAKAAEAAKAAEAAKAAADAKKAAAE 240
 RESULT 8
 Q82GZ2 PRELIMINARY; PRT; 388 AA.
 AC Q82GZ2;
 DT 01-MAR-2002 (TReMBLrel. 20, Created)
 DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
 DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)

```

DE TOLA colicin import membrane protein.
GN TOLA OR YP01123.
OS Yersinia pestis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Yersinia.
NCBI_Taxid=632;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CO-92 / Biovar Orientalis;
RX MEDLINE=21470413; PubMed=11586360;
RA Parhill J., Wren B.W., Thomson N.R., Tibball R.W., Holden M.T.G.,
RA Prentice M.B., Sebahia M., James K.D., Churcher C., Mungall K.L.,
RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdeno-Tarraga A.M.,
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA Felwell T., Hamlin N., Holroyd S., Jegels K., Karlyshev A.V.,
RA Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,
RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrett B.G.;
RA "Genome sequence of Yersinia pestis, the causative agent of plague.",
RL Nature 413:523-527(2001).
DR EMBL; AJ414146; CAC89966.1; -.
KW Complete proteome.
SQ SEQUENCE 388 AA; 40424 MW; 81447B04B30A7E7C CRC64;

Query Match 34.3%; Score 178; DB 16; Length 388;
Best Local Similarity 52.1%; Pred. No. 4.6e-05;
Matches 63; Conservative 12; Mismatches 28; Indels 18; Gaps 7;

QY 2 KYAKKAEYKAAKAAKAAK-----AYAKKEA-KAYKA-AEAKKAAKAAKAYKAA 53
DB 140 KQAEQOKIAAAAVAKAKKEQKQAEFTAAQAKTEAKTKYKQAEQAKAEAKKEAAVA 199
QY 54 AKAKKAEYKAAKAAKAAK-----AEKKEVAAAEAKKAAKAAK-AVYAAKAAKAAKAA 105
DB 200 AAKAKQA-DADAKKAVEAEKAAADAEKKAADAE-KKAAAKKVAATAAEAKKAAKAA 257
QY 106 A 106
DB 258 A 258

RESULT 9
ID Q8CZ28 PRELIMINARY; PRT; 393 AA.
AC Q8CZ28;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Membrane spanning protein.
GN TOLA OR Y1056.
OS Yersinia pestis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Yersinia.
NCBI_Taxid=632;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=KIMS / Biovar Mediaevalis;
RX MEDLINE=21237863; PubMed=12142430;
RA Deng W., Burland V., Plunkett G. III, Boutin A., Mayhew G.F., Lies P.,
RA Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C.,
RA Fetherston J.D., Lindler L.B., Brubaker R.R., Plano G.V.,
RA Straley S.C., McDonough K.A., Nilles M.L., Watson J.S., Blattner F.R.,
RA Perry R.D.;
RA "Genome sequence of Yersinia pestis KIM.",
RL J. Bacteriol. 184:4601-4611(2002).
DR EMBL; AB013906; AAM86606.1; -.
SQ SEQUENCE 393 AA; 41012 MW; 1E3E4F87B533481 CRC64;

Query Match 34.3%; Score 178; DB 16; Length 393;
Best Local Similarity 52.1%; Pred. No. 4.7e-05;
Matches 63; Conservative 12; Mismatches 28; Indels 18; Gaps 7;

QY 2 KYAKKAEYKAAKAAKAAK-----AYAKKEA-KAYKA-AEAKKAAKAAKAYKAA 53

```

```

DB 145 KQAEQOKIAAAAVAKAKKEQKQAEFTAAQAKTEADTKVKAQAEQKAEAEAKKEAAVA 204
QY 54 AKAKKAEYKAAKAAKAAK-----AEKKEVAAAEAKKAAKAAK-AVYAAKAAKAAKAA 105
DB 205 AAKAKQA-DADAKKAVEAEKAAADAEKKAADAE-KKAAAKKVAATAAEAKKAAKAA 262
QY 106 A 106
DB 263 A 263

RESULT 10
ID Q8XVNT PRELIMINARY; PRT; 200 AA.
AC Q8XVNT;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Probable histone H1 protein.
GN RSC2793 OR RS00453.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Ralstoniaceae; Ralstonia.
NCBI_Taxid=305;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GM11000;
RX MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Arlat M., Billault A., Broctier P., Gams J.C., Cattoilco L.,
RA Chandler M., Choisme N., Claudel-Renard C., Cunac S., Demange N.,
RA Gaspier C., Lavie M., Moisan A., Robert C., Sautin W., Schlex T.,
RA Weisenbach J., Boucher C.A.;
RA "Genome sequence of the plant pathogen Ralstonia solanacearum.",
RL Nature 415:497-502(2002).
DR EMBL; AL646071; CAD16500.1; -.
DR InterPro; IPR005819; Histone_H5.
DR PRINTS; PR00624; HISTONEH5.
KW Complete proteome.
SQ SEQUENCE 200 AA; 19279 MW; D3831B590510272D CRC64;

Query Match 34.0%; Score 176.5; DB 16; Length 200;
Best Local Similarity 54.5%; Pred. No. 3.2e-05;
Matches 67; Conservative 5; Mismatches 32; Indels 19; Gaps 8;

QY 1 AKVYAKKAEYKAAKAAK-----KKAYAKKEAKYKAAKAAK-----KAEAKTY- 49
DB 42 AKKVA-AKVAAKQAAPAKKAAVGVAAKKAAPAKKAIVKVAAPAKKAAVKKVA 99
QY 50 AKEAAKAKKAEYKAAKAAKAAK-----AEKKEVAAAEAKKAAKAAKAAKAAKAA 106
DB 100 AKKAAAPAKKAAVAVKVAAKKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAA 156
QY 107 YEA 109
DB 157 KKA 159

RESULT 11
ID Q9CM70 PRELIMINARY; PRT; 389 AA.
AC Q9CM70;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE TOLA OR PM0968.
GN Pasteurella multocida.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Pasteurella.
NCBI_Taxid=747;
RN [1]

```

```
RP SEQUENCE FROM N.A.
RC STRAIN=PM70;
RA MEDLINE=21145866; PubMed=11248100;
RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
RT "Complete genomic sequence of Pasteurella multocida PM70."
RT Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
DR EMBL; AB06136; AAK03052.1; -.
DR HSSP; P19934; 1TOL.
DR InterPro; IPR000533; Tropomyosin.
DR PRINTS; PR00194; TROPOMYOSIN.
DR Complete proteome.
SQ SEQUENCE 389 AA; 42197 MW; B4032F2A2FD9E94B CRC64;

Query Match
33.4%; Score 173.5; DB 16; Length 389;
Best Local Similarity 48.7%; Pred. No. 9.4e-05;
Matches 55; Conservative 18; Mismatches 31; Indels 9; Gaps 4;

OY 6 KKAERKAYAKK-AKAAER-KAYAKKAYKAAAKKKAKA--EAKKYAKA-----AKA 56
DB 145 KQAEKAKQALAEAKLAKAEAKKRLAALAKQAEBAKKAEBEAKKAEKAKAEAKA 204
OY 57 KKEAYKAEKKYAKAKAEKKEVAAAEAKKAEAAKAYKAAKAAKAEAAVEA 109
DB 205 KVEKAKAEAEKAYKAEKAKAEAAKAEKAKAEKAKAEKAKADAEAAQRA 257

RESULT 12
061164 PRELIMINARY; PRT; 1701 AA.
AC 061164;
DT 01-AUG-1998 (TRENBLREL. 07, Created)
DT 01-AUG-1998 (TRENBLREL. 07, Last sequence update)
DT 01-OCT-2002 (TRENBLREL. 22, Last annotation update)
DE Erythrocyte binding protein.
GN MAEBL.
OS Plasmodium yoelii yoelii.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_Taxid=73239;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=YM;
RX MEDLINE=98115903; PubMed=9448314;
RA Kape S.H.I., Noe A.R., Fraser T.S., Blair P.L., Adams J.H.;
RT "A family of chimeric erythrocyte binding proteins of malaria parasites."
RT Proc. Natl. Acad. Sci. U.S.A. 95:1230-1235(1998).
RL DR EMBL; AF03186; AAC05366.1; -.
SQ SEQUENCE 1701 AA; 199268 MW; EDA8E2DEFD87C8BA CRC64;

Query Match
32.9%; Score 170.5; DB 5; Length 1701;
Best Local Similarity 53.5%; Pred. No. 0.00058;
Matches 54; Conservative 11; Mismatches 31; Indels 5; Gaps 5;

OY 5 AKKAERKAYAKKAEKKAERKAYAKAEAKKKAKAEKKAERKAYKAEKKAERKAYK- 63
DB 1205 AKKAEE-RKKAERK-KAEAKKAEKKAERK-KAEAKKALEKKKSEAKKAL 1261
OY 64 EAKKYAKAKAEKKEVAAAEAKKAEAAKAYKAAKAAKAE 104
DB 1262 ERKKAERKAEKKAERKAEKKAERKAEKKAERKAEKKAERKAEKKAERKAE 1301

RESULT 13
039576 PRELIMINARY; PRT; 232 AA.
AC 039576;
DT 01-NOV-1996 (TRENBLREL. 01, Created)
DT 01-NOV-1996 (TRENBLREL. 01, Last sequence update)
DT 01-MAR-2003 (TRENBLREL. 23, Last annotation update)
DE Histone H1.
GN CH1.
OS Chlamydomonas reinhardtii.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
```

```
OC Chlamydomonadaceae; Chlamydomonas.
OX NCBI_Taxid=3055;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=96120862; PubMed=8590479;
RA Fabry S., Muller K., Lindner A., Park P.B., Cornelius T., Schmitt R.;
RT "The organization structure and regulatory elements of Chlamydomonas histone genes reveal features linking plant and animal genes."
RT Curr. Genet. 28:333-345(1995).
RL DR EMBL; U16726; AAA98452.1; -.
DR HSSP; P02259; IHST.
DR InterPro; IPR005818; Histone H1/H5.
DR InterPro; IPR005819; Histone H5.
DR InterPro; IPR003216; Linkerhist N.
DR Pfam; PF00538; Linkerhistone_I.
DR PRINTS; PR00624; HISTONEH5.
DR ProDom; PD000373; Linkerhist_N; 1.
DR SMART; SM00526; H15; 1.
SQ SEQUENCE 232 AA; 24693 MW; 2D006AE44A8FA037 CRC64;

Query Match
32.8%; Score 170; DB 10; Length 232;
Best Local Similarity 46.3%; Pred. No. 0.0001;
Matches 57; Conservative 10; Mismatches 38; Indels 18; Gaps 5;

OY 2 KKYAKKA---EKAYKAKAKAEKKAERKAEKKAERKAEKKAERKAEKKAERKAEK- 53
DB 100 KKAERKAAKTPKKAAPKKEGAVKTKAPKKEGKPKAKAEKKEKPKKEGKKAAPK 159
OY 54 ---AKAKERKAYKAEKKAERKAEKKEVAAAE---AKKAERK---AYKAAKAAK 103
DB 160 AEKKRKAAPKAPKPTTKAAAKPKAEKPKKAAKPKAEKPKKAAKPKAEKKAAPK 219
OY 104 EAA 106
DB 220 KSA 222

RESULT 14
001395 PRELIMINARY; PRT; 275 AA.
AC 001395;
DT 01-JUL-1997 (TRENBLREL. 04, Created)
DT 01-JUL-1997 (TRENBLREL. 04, Last sequence update)
DT 01-OCT-2002 (TRENBLREL. 22, Last annotation update)
DE Axoneme-associated protein MST101(3).
GN MST101(3) OR DHMST101.
OS Drosophila hydei (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_Taxid=7224;
RN (1)
RP SEQUENCE FROM N.A.
RA Neesen J., Heinlein U.A.O., Buenemann H.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: POSSIBLE STRUCTURAL ROLE IN THE SPERM TAIL (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC -1- TISSUE SPECIFICITY: TESTIS (BY SIMILARITY).
CC -1- DOMAIN: THE PREDOMINANT STRUCTURE IS ALPHA-HELICAL.
DR EMBL; U85627; AAB51369.1; -.
DR Flybase; FBgn0020732; Dhyd1mst101(3).
DR KWSperm; Repeat; Multigene family.
KW DOMAIN
FT 64
FT 255
SQ SEQUENCE 275 AA; 30436 MW; 76BAA7B2A2DF732C CRC64;

Query Match
32.7%; Score 169.5; DB 5; Length 275;
Best Local Similarity 48.2%; Pred. No. 0.00013;
Matches 54; Conservative 16; Mismatches 29; Indels 13; Gaps 6;

OY 1 AKYAKKAEKAYAKK--AKAAEKKAYAKKAEKAYKAAEAKKKAK-ASAERYAKAEAAVAK 57
```

```

Db      51 AEDVKKCEBAAKKCAEAAKKEAAEK-----KCAEAAKKEAAEKKCAEAAKKE 106
Qy      58 KEAYKAAKKAAYKAAKKEAYAAAEAKK-AEAAKAYKAAKAAKAAKAAAYE 108
Db      107 QEA--AQKKCAELAKEKE--AAKKCAEAAKKEAAEKKEAAEKCEAAFK 153

```

RESULT 15

```

039598      PRELIMINARY;      PRT;      265 AA.
AC      039598;
DT      01-NOV-1996 (TrEMBLrel. 01, Created)
DT      01-NOV-1996 (TrEMBLrel. 01, last sequence update)
DT      01-OCT-2002 (TrEMBLrel. 22, last annotation update)
DE      CGCR-4 product (Fragment).
GN      CGCR-4.
OS      Chlamydomonas reinhardtii.
OC      Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC      Chlamydomonadaceae; Chlamydomonas.
OX      NCBI_TaxID=3055;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=92119224; PubMed=1731966;
RA      Wakarchuk W.W., Muller F.W., Beck C.F.;
RT      "Two GC-rich DNA elements of Chlamydomonas reinhardtii with complex
RT      arrangements of directly repeated sequence motifs."
RL      Plant Mol. Biol. 18:143-146(1992).
DR      EMBL; X17208; CAA35080.1; -.
FT      NON_TER
SQ      SEQUENCE      265 AA; 26216 MW; B3518B7377CF782 CRC64;

```

```

Query Match      31.6%; Score 164; DB 10; Length 265;
Best Local Similarity 39.4%; Pred. No. 0.00029;
Matches 50; Conservative 21; Mismatches 38; Indels 18; Gaps 2;

```

```

Qy      1 AKKYAKKAKEA-----YAKKAKAAKKEKAYAKKEAKYKAAEAKKKAAEA 46
Db      79 AEAKEADEAEAPAAEAAPAAVAEAWAAAEAEAPAAEAAPAAAEVAPAAAOA 138
Qy      47 KKYAKAAKKAKEAYKAAKKAAYKAAK-----AEKKEYAAAEAKKAEAKAYKAAKAAA 102
Db      139 RVAAEAPAAAEAPAAAEAPAAAEAPAAAEAPAAAEAPAAAEAPAAAEAPAA 198
Qy      103 KEAAAYEA 109
Db      199 EAKAKEA 205

```

Search completed: January 21, 2004, 09:00:27
 Job time : 44.8831 secs

PA	(TEVA-)-TEVA PHARM USA INC.
XX	Gad A, IIs D;
XX	WPI; 2000-317499/27.
DR	Copolymer 1 related polypeptides used as molecular weight markers for
PT	glatiramer acetate and for treatment and prevention of immune diseases
PS	Claim 10; Page 14; 72pp; English.
XX	AAY82571 to AAY82577 represent specifically claimed copolymer molecular
CC	weight TV-marker polypeptides from the present invention. The present
CC	invention describes polypeptides (I) for determining the molecular
CC	weight of a copolymer (CP), which has an identified molecular weight
CC	and an amino acid composition corresponding to the copolymer. The
CC	polypeptides of the invention are used as molecular weight markers for
CC	glatiramer acetate related tetrapolymers. The polypeptides may also be
CC	used for treating and preventing immune diseases in a mammal. Autoimmune
CC	diseases which may be treated include either cell-mediated or
CC	antibody-mediated diseases. Such diseases include arthritic conditions,
CC	demylinating diseases and inflammatory conditions, e.g. multiple
CC	sclerosis, rheumatoid arthritis, osteoarthritis, autoimmune haemolytic
CC	anaemia, autoimmune coporitis, autoimmune thyroiditis, autoimmune
CC	uveoretinitis, Crohn's disease, chronic immune thrombocytopenia
CC	purpure, colitis, contact sensitivity disease, diabetes mellitus, Graves
CC	disease, Guillain-Barre's syndrome, Hashimoto's disease, idiopathic
CC	myxoedema, myasthenia gravis, psoriasis, pemphigus vulgaris, or systemic
CC	lupus erythematosus. Mediated diseases which can be treated
CC	include host-versus-graft disease, graft-versus-host disease, and
CC	delayed-type hypersensitivity. The polypeptides of the invention have
CC	defined molecular weights and physical properties which are analogous to
CC	glatiramer acetate molecules, which makes them ideal for use as
CC	molecular weight markers.
SQ	Sequence 109 AA:
Query Match	100.0%; Score 519; DB 21; Length 109;
Best Local Similarity	100.0%; Pred. No. 1.5e-36;
Matches 109; Conservative	0; Mismatches 0; Indels 0; Gaps 0
OY	1 AKKAKKAKEKKAKAEEKKAAKYAAAEKKAQAARAKKATKAARKAAAKKXA 60
DB	1 AKKAKKAKEKKAKAEEKKAAKYAAAEKKAQAARAKKATKAARKAAAKKXA 60
OY	61 YKAEKKTKAKAKAKEEKKYAABAKKAERAKAYKKAEBAKAAKAAEAAYA 109
DB	61 YKAEKKTKAKAKAKEEKKYAABAKKAERAKAYKKAEBAKAAKAAEAAYA 109
RESULT 2	
ID	AAY82576
AC	AAY82576 standard; peptide; 86 AA.
XX	AAY82576;
DT	
XX	28-JUL-2000 (first entry)
DE	Copolymer molecular weight TV-marker amino acid sequence SEQ ID NO:6.
XX	Copolymer; molecular weight marker; TV-marker; immune disease;
KM	gliatimer acetate; autoimmune disease; antiarthritic; neuroprotective;
KM	osteoprotic; immunosuppressive; antithyroid; antiinflammatory;
KM	antiadrenergic; thyromimetic; haemostatic; antipsoriatic; dermatological;
KM	inflamamentary condition; multiple sclerosis; rheumatoid arthritis;
KM	Crohn's disease; chronic immune thrombocytopoenia purpura; colitis;
KM	diabetes mellitus; Graves disease; Guillain-Barre's syndrome; psoriasis;
KM	Hashimoto's disease; idiopathic myxoedema; myasthenia gravis;
KM	pemphigus vulgaris; systemic lupus erythematosus.
XX	Unidentified.
OS	

MO200018794-A1.

XX PD 06-APR-2000.

XX PF 24-SEP-1999; 99WO-US22402.

XX PR 25-SEP-1998; 98US-0101693.

PA (YEDA) YEDA RES & DEV CO LTD.
PA (TEVA-) TEVA PHARM USA INC.

P1 Gad A, Lis D;

PS WPI: 2000-317499/27.

XX Copolymer 1 related polypeptides used as molecular weight markers for
PT glutarimer acetate and for treatment and prevention of immune diseases
DR

XX Claim 10; Page 14; 72pp; English.

XX AY82571 to AY82577 represent specifically claimed copolymer molecular
CC weight TV-marker polypeptides from the present invention. The present
CC invention describes polypeptides (I) for determining the molecular
CC weight of a copolymer (CP), which has an identified molecular weight
CC and an amino acid composition corresponding to the copolymer. The
CC polypeptides of the invention are used as molecular weight markers for
CC glutarimer acetate related tetrapolymers. The polypeptides may also be
CC used for treating and preventing immune diseases in a mammal. Autoimmune
CC diseases which may be treated include either cell-mediated or
CC antibody-mediated diseases. Such diseases include arthritic conditions,
CC demyelinating diseases and inflammatory conditions, e.g. multiple
CC sclerosis, rheumatoid arthritis, osteoarthritis, autoimmune haemolytic
CC anaemia, autoimmune oophoritis, autoimmune thyroiditis, autoimmune
CC ulcerative colitis, Crohn's disease, chronic immune thrombocytopenia
CC purpura, contact sensitivity disease, diabetes mellitus, Graves
CC disease, Guillain-Barre's syndrome, Hashimoto's disease, idiopathic
CC myxoedema, myasthenia gravis, psoriasis, pemphigus vulgaris, or systemic
CC lupus erythematosus. Mediated-mediated diseases which can be treated
CC include host-versus-graft disease, graft-versus-host disease, and
CC delayed-type hypersensitivity. The polypeptides of the invention have
CC defined molecular weights and physical properties which are analogous to
CC glutarimer acetate molecules, which makes them ideal for use as
CC molecular weight markers.

XX SQ Sequence 86 AA;

QY Query Match 60.2%; Score 312.5; DB 21; Length 86;
Best Local Similarity 72.1%; Pred. No. 2.2e-19;
Matches 80; Conservative 1; Mismatches 3; Indels 27; Gaps 5

DB 1 AKVAKKAKEAYKAKKAEKKAVAYKAAEKKAKKAAEKYKAAEAARAKKEA 60
1 AKKVAKK-EKAYAKKA-----EKAKKKEAAAYKAAEAKKKA----- 36

QY 61 YKAEKKYKAKKAEKKKYAAAEK-KAEA-KAYKAAAYKAAEAAYEA 109
| | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | |
37 -KAEKKYKAKKAEKKYEAAEAYKAAKGAAYKAAAYKAAEAAYEA 86

ID RESULT 3
AY82575
AY82575 standard; peptide; 77 AA.

AC AY82575;

XX DT 28-JUL-2000 (first entry)

XX Copolymer molecular weight TV-marker amino acid sequence SEQ ID NO.5.

XX Copolymer; molecular weight marker; TV-marker; immune disease;
KW glutarimer acetate; autoimmune disease; antiarthritic; neuroprotective;
KW osteoporitic; immunosuppressive; antihypertensive; antiinflammatory;
KW antidiabetic; thyromimetic; hemostatic; antipsoriatic; dermatological;

OS Unidentified.
XX
XX WO200018794-A1.
PN
XX PD 06-APR-2000.
XX
XX 24-SEP-1999; 99WO-US22402.
XX
XX 25-SEP-1998; 98US-0101693.
XX
XX (YEDA) YEDA RES & DEV CO LTD.
PA (TEVA-) TEVA PHARM USA INC.
XX
XX Gad A, Lis D;
P1
XX MPI; 2000-317499/27.
DR
XX
PT Copolymer 1 related polypeptides used as molecular weight markers for
FT glatiramer acetate and for treatment and prevention of immune diseases
XX
PS Claim 10; Page 14; 72pp; English.

AY82571 to AY82577 represent specifically claimed copolymer molecular weight TV-marker polypeptides from the present invention. The present invention describes polypeptides (I) for determining the molecular weight of a copolymer (CP), which has an identified molecular weight and an amino acid composition corresponding to the copolymer. The polypeptides of the invention are used as molecular weight markers for glatiramer acetate related tetrapolymers. The polypeptides may also be used for treating and preventing immune diseases in a mammal. Autoimmune diseases which may be treated include either cell-mediated or antibody-mediated diseases. Such diseases include arthritic conditions, demyelinating diseases and inflammatory conditions, e.g. multiple sclerosis, rheumatoid arthritis, osteoarthritis, autoimmune haemolytic anaemia, autoimmune oophoritis, autoimmune thyroiditis, autoimmune puerperalitis, Crohn's disease, chronic immune thrombocytopenia purpura, colitis, contact sensitivity disease, diabetes mellitus, Graves disease, Guillain-Barre's syndrome, Hashimoto's disease, idiopathic myxoedema, myasthenia gravis, psoriasis, pemphigus vulgaris, or systemic lupus erythematosus. Mediated-mediated diseases which can be treated include host-versus-graft disease, graft-versus-host disease, and delayed-type hypersensitivity. The polypeptides of the invention have defined molecular weights and physical properties which are analogous to glatiramer acetate molecules, which makes them ideal for use as molecular weight markers.

Sequence 77 AA;

Query Match 55.7%; Score 289; DB 21; Length 77;
Best Local Similarity 67.0%; Pred. No. 1.8e-17;
Matches 73; Conservative 1; Mismatches 3; Indels 32; Gaps 4,

OY 1 AKTAKKAEKAYAKKAKEKKAYAKKEBAKYAKAEEKAKKAAKAAKAAKAAEAYEA 109
DB 1 AKKVAKK-EKVYAKKA-----EKAKKKAEKAYAKKAEEKKA----- 36

OY 61 YKAEAKTKAKAKAKEKKEYYAAAEAKKAEEAKYKKAEEAAKAAKAAEAYEA 109
DB 37 -KAEAKTKAKAKAKEKKEYYAAAEAK-----YKAEAKKAAKAAEAYEA 77

RESULT 4
AY82574
ID AY82574 standard; peptide; 66 AA.
XX
AC AY82574;

DT 28-JUL-2000 (first entry)

XX Copolymer molecular weight TV-marker amino acid sequence SEQ ID NO:4.

XX

XX Copolymer: molecular weight marker; TV-marker; immune disease;

XX glatiramer acetate; autoimmune disease; antiarthritic; neuroprotective;

XX ocreoapathic; immunosuppressive; antithyroid; antiinflammatory;

XX antidiabetic; thyromimetic; haemostatic; antipsoriatic; dermatological;

XX antianaemic; immunosuppressive; demyelinating disease; osteoarthritis;

XX inflammatory condition; multiple sclerosis; rheumatoid arthritis;

XX Crohn's disease; chronic immune thrombocytopaenia purpura; colitis;

XX diabetes mellitus; Graves disease; Guillain-Barre's syndrome; psoriasis;

XX Hashimoto's disease; idiopathic myxoedema; myasthenia gravis;

XX pemphigus vulgaris; systemic lupus erythematosus.

XX

XX Unidentified.

XX

XX WO200018794-A1.

XX

XX 06-APR-2000.

XX

XX 24-SEP-1999; 99WO-US22402.

XX

XX 25-SEP-1998; 98US-0101693.

XX

XX (VEDA) VEDA RES & DEV CO LTD.

XX (TEVA-) TEVA PHARM USA INC.

XX

XX Gad A, Lis D;

XX

XX WPI; 2000-317499/27.

DR

XX Copolymer 1 related polypeptides used as molecular weight markers for

XX glatiramer acetate and for treatment and prevention of immune diseases

PT

XX

XX Claim 10; Page 14; 72pp: English.

PS

XX AAY82571 to AAY82577 represent specifically claimed copolymer molecular

XX weight TV-marker polypeptides from the present invention. The present

XX invention describes polypeptides (I) for determining the molecular

XX weight of a copolymer (Cp), which has an identified molecular weight

XX and an amino acid composition corresponding to the copolymer. The

XX polypeptides of the invention are used as molecular weight markers for

XX glatiramer acetate related tetrapolymers. The polypeptides may also be

XX used for treating and preventing immune diseases in a mammal. Autoimmune

XX diseases which may be treated include either cell-mediated or

XX antibody-mediated diseases. Such diseases include arthritic conditions,

XX demyelinating diseases and inflammatory conditions, e.g. multiple

XX sclerosis, rheumatoid arthritis, osteoarthritis, autoimmune haemolytic

XX anaemia, autoimmune oophoritis, autoimmune thyroiditis, autoimmune

XX uveoretinitis, Crohn's disease, chronic immune thrombocytopaenia

XX purpura, colitis, contact sensitivity disease, diabetes mellitus, Graves

XX disease, Guillain-Barre's syndrome, Hashimoto's disease, idiopathic

XX myxoedema, myasthenia gravis, psoriasis, pemphigus vulgaris, or systemic

XX lupus erythematosus. Mediated-mediated diseases which can be treated

XX include host-versus-graft disease, graft-versus-host disease, and

XX delayed-type hypersensitivity. The polypeptides of the invention have

XX defined molecular weights and physical properties which are analogous to

XX glatiramer acetate molecules, which makes them ideal for use as

XX molecular weight markers.

XX

XX

XX Sequence 66 AA;

XX

XX Query Match 44.0%; Score 228.5; DB 21; Length 66;

XX Best Local Similarity 56.9%; Pred. No. 1.8e-12;

XX Matches 62; Conservative 0; Mismatches 4; Indels 43; Gaps 4;

XX

XX 1 AKYAKKAEEKAYAKKAAEKKAYAKKAAAYRAAEAKKKAKAEKKYAKAEAKKAKKA 60

XX ||||| ||||| || ||||| ||

XX 1 AKYAKK- EKAYAKKAKA-----EAKGAKKA----- 25

XX

XX 61 YKAEKKYAKAKAEKKEYYAAAEKKAEAAAYAYRAEAKKAAAKKAAEAAYEA 109

Db 26 -KAEAKYAKAKAEKKEYPAAAEAK-----YKAEAAAKAEKAEYA 66

RESULT 5

ID AAR06445 standard; protein; 154 AA.

AC AAR06445;

DT 25-MAR-2003 (updated)
DT 03-JAN-1991 (first entry)

DE Recombinant copolymer 1-77, myelin basic protein analogue.

KW Recombinant copolymer 1; COP-1-77; myelin basic protein; MBP;
KW immunological activity; autoimmune encephalomyelitis;
KW multiple sclerosis;

OS Synthetic.

PN EP383620-A.

PD 22-AUG-1990.

PF 16-FEB-1990; 90EP-0301700.

PR 07-FEB-1990; 90US-0473845.

PR 17-FEB-1989; 89US-0312541.

PA (REPK) REPLIGEN CORP.

PI Cook KS;

DR MPI; 1990-255848/34.

DR N-PSDB; AAQ05664.

PT Producing genes encoding random polymers of aminoacid(s) - for
PT producing recombinant polypeptide(s) with biological and/or
PT immunological activity

PS Disclosure; Fig 11; 25pp; English.

XX To improve the expression of rCOP-1 polypeptides in E. coli, genes

CC coding for rCOP-1-77 were subcloned from PREV 2.1 to pBG3-2deltan

CC (deposit: 20-NOV-1984 US4691009, NRL B-15910), a plasmid used to

CC express Protein A. The resulting plasmids encode fusion proteins

CC consisting of beta-glucuronidase, Protein A, and rCOP-1 sequences.

CC A methionine residue occurs between the Protein A and rCOP-1

CC sequences, originating from the 5' linker sequence, in order that

CC the COP-1 polypeptide may be cleaved from the fusion protein.

CC rCOP-1-77 contains oligonucleotide duplexes incooding the following

CC segments: YKK, EAR, KAK, AAK, and AAA. The N-terminal alanine residue

CC is left behind following CNBR cleavage of the fusion protein.

CC The product prevents or arrests experimental autoimmune

CC encephalomyelitis. They are used to prevent, arrest or control a

CC demyelinating disorder, e.g. multiple sclerosis. They may also

CC be used as additives to hair care products to confer beneficial

CC effects on damaged hair or as supplements for diets deficient in

CC certain amino acids.

CC See also AAQ05665.

CC (Updated on 25-MAR-2003 to correct PA field.)

CC Sequence 154 AA;

CC Query Match 40.3%; Score 209; DB 11; Length 154;

CC Best Local Similarity 53.7%; Pred. No. 1.9e-10;

CC Matches 72; Conservative 8; Mismatches 20; Indels 34; Gaps 9;

CC

CC

CC

CC

CC

CC

CC

OY 45 EAKYKAEAKA--KKEAYKAEAKYKAKAEK-----KEAYAEAKKAEAA--K 91
DB 82 EYKKAKEAAAEYKKEAEAEYKKYKKAKEAKYKKAKEAEKAAAEKAEAEYK 141

OY 92 AYKAEAAKAAKAE 105

DB 142 KYKKEAEK--AKEA 153

RESULT 6

ID AAR06446 standard; protein; 106 AA.

AC AAR06446;

DT 25-MAR-2003 (updated)
DT 03-JAN-1991 (first entry)

DE Recombinant copolymer 1-19, myelin basic protein analogue.

KW Recombinant copolymer 1; COP-1-19; myelin basic protein; MBP;
KW immunological activity; autoimmune encephalomyelitis;
KW multiple sclerosis;

OS Synthetic.

PN EP383620-A.

PD 22-AUG-1990.

PF 16-FEB-1990; 90EP-0301700.

PR 07-FEB-1990; 90US-0473845.

PR 17-FEB-1989; 89US-0312541.

PA (REPK) REPLIGEN CORP.

PI Cook KS;

DR MPI; 1990-255848/34.

DR N-PSDB; AAQ06446.

PT Producing genes encoding random polymers of aminoacid(s) - for
PT producing recombinant polypeptide(s) with biological and/or
PT immunological activity

PS Disclosure; Fig 12; 25pp; English.

XX To improve the expression of rCOP-1 polypeptides in E. coli, genes

CC coding for rCOP-1-19 were subcloned from PREV 2.1 to pBG3-2deltan

CC (deposit: 20-NOV-1984 US4691009, NRL B-15910), a plasmid used to

CC express Protein A. The resulting plasmids encode fusion proteins

CC consisting of beta-glucuronidase, Protein A, and rCOP-1 sequences.

CC A methionine residue occurs between the Protein A and rCOP-1

CC sequences, originating from the 5' linker sequence, in order that

CC the COP-1 polypeptide may be cleaved from the fusion protein.

CC rCOP-1-19 contains oligonucleotide duplexes incooding the following

CC segments: YKK, AAB, KAK, EKA, KKA, YEA, AKA KEA, and AAA. The

CC N-terminal alanine residue is left behind following CNBR cleavage of the

CC fusion protein.

CC The product prevents or arrests experimental autoimmune

CC encephalomyelitis. They are used to prevent, arrest or control a

CC demyelinating disorder, e.g. multiple sclerosis. They may also

CC be used as additives to hair care products to confer beneficial

CC effects on damaged hair or as supplements for diets deficient in

CC certain amino acids.

CC See also AAQ05664.

CC (Updated on 25-MAR-2003 to correct PA field.)

CC Sequence 106 AA;

CC Query Match 36.0%; Score 187; DB 11; Length 106;

CC Best Local Similarity 55.0%; Pred. No. 8.7e-09;

CC

CC

CC

CC

CC

Matches 60; Conservative 11; Mismatches 20; Indels 18; Gaps 6

Qy 5 AKKAEKAVAKKAKAEKKAVAKKEAKAVAKAAEAKKAKAEAKKYAKAEAAKAEVAYAAE 64

Db 1 AKAEEKAKAAK-KAYEAEKAKAVKYEAK--KAEKAEKAEKAAAEKKAKAEKAEK----- 51

Qy 65 AKKAKAKAKAEKKEVAAAEAKKAEAA---KAYKAEAAKAAK-AAAE 108

Db 52 AKEAEKAEKAEK-----AKEAKKAKAAEAKAEKAEKAKAEKAAEKAAKAAK 96

RESULT 7

ID AAY82573

AY82573

AC AAY82573; peptide; 56 AA.

XX AAY82573;

DT 28-JUL-2000 (first entry)

DE Copolymer molecular weight TV-marker amino acid sequence SEQ ID NO:3.

XX Copolymer: molecular weight marker; TV-marker; immune diseases;

KW glatiramer acetate; autoimmune disease; antiarthritis; neuroprotective;

KW oseropathic; immunosuppressive; antihypoid; antinflammatory;

KW antidiabetic; thyromimetic; haemostatic; antiporistic; dermatological;

KW antianemic; immunosuppressive; demyelinating disease; osteoarthritis;

KW inflammatory condition; multiple sclerosis; rheumatoid arthritis;

KW Crohn's disease; chronic immune thrombocytopaenia purpura; colitis;

KW diabetes mellitus; Graves disease; Guillain-Barre's syndrome; psoriasis;

KW Hashimoto's disease; idiopathic myxoedema; myasthenia gravis;

KM pemphigus vulgaris; systemic lupus erythematosus.

XX unidentified.

XX

OS

XX

PN WO200018794-A1.

XX

PD 06-APR-2000.

XX

XX 24-SEP-1999; 99WO-US22402.

PF

XX 25-SEP-1998; 98US-0101693.

PR

XX (YEDA) YEDA RES & DEV CO LTD.

PA (TEVA-) TEVA PHARM USA INC.

XX

XX Gad A, Lis D;

PI

DR WPI; 2000-317499/27.

XX

PT Copolymer 1 related polypeptides used as molecular weight markers for

PT glatiramer acetate and for treatment and prevention of immune diseases

XX

XX Claim 10; Page 14; 72pp; English.

XX

XX AAY82571 to AAY82577 represent specifically claimed copolymer molecular

CC weight TV-marker polypeptides from the present invention. The present

CC invention describes polypeptides (I) for determining the molecular

CC weight of a copolymer (CP), which has an identified molecular weight

CC and an amino acid composition corresponding to the copolymer. The

CC polypeptides of the invention are used as molecular weight markers for

CC glatiramer acetate related tetrapolymers. The polypeptides may also be

CC used for treating and preventing immune diseases in a mammal. Autoimmune

CC diseases which may be treated include either cell-mediated or

CC antibody-mediated diseases. Such diseases include arthritic conditions,

CC demyelinating diseases and inflammatory conditions, e.g. multiple

CC sclerosis, rheumatoid arthritis, osteoarthritis, autoimmune haemolytic

CC anaemia, autoimmune oophoritis, autoimmune thyroiditis, autoimmune

CC uveoretinitis, Crohn's disease, chronic immune thrombocytopaenia

CC purpura, colitis, contact sensitivity diseases, diabetes mellitus, Graves

CC disease, Guillain-Barre's syndrome, Hashimoto's disease, idiopathic

CC myxoedema, myasthenia gravis, psoriasis, pemphigus vulgaris, or systemic

CC lupus erythematosus. Mediated-mediated diseases which can be treated

CC include host-versus-graft disease, graft-versus-host disease, and

```

CC delayed-type hypersensitivity. The polypeptides of the invention have
CC defined molecular weights and physical properties which are analogous to
CC gliatramer acetate molecules, which makes them ideal for use as
CC molecular weight markers.
XX
XX
SQ Sequence 56 AA;
XX
XX
Query Match 34.8%; Score 180.5; DB 21; Length 56;
Best Local Similarity 45.9%; Pred. No. 1.6e-08;
Matches 50; Conservative 3; Mismatches 3; Indels 53; Gaps 3;
OY
1 AKKTKAKKAKKAYKAKKAAKKEKKAYKAKKAYKAAKAKKKAKEAKKAYKAAKAKKEA 60
Db 1 AKKTKAKK- EKKYAKKA----- EKKAKKAKEAKATYAAAKKKAKKEK----- 39
OY 61 YKAEAKTKYAKAAKAKKEKYEAAAEKKAEEAAKYEKAEAAKAAKEAYEA 109
Db 40 -----YKAEAKKAAKAAKEAYEA 56
XX
XX
RESULT 8
AA98499
ID AA98499 standard; Peptide; 100 AA.
XX
XX AA98499;
AC
XX
XX 31-JUL-2000 (first entry)
DE Peptide #10 used in nucleic acid transporter system.
XX
XX Transporter system; nucleic acid delivery; gene therapy; cancer;
XX cardinogenesis; cardiovascular disease; infection.
XX
XX Synthetic.
OS
XX
XX US6033884-A.
PN
XX
XX 07-MAR-2000.
PD
XX
XX 14-DEC-1993; 93US-0167641.
PF
XX
XX 20-MAR-1992; 92US-0855389.
PR 19-MAR-1993; 93WO-US027225.
XX
XX
PA (BAYU ) BAYLOR COLLEGE MEDICINE.
XX
XX
PI Gottchalk S, Sparrow J, Cristiano RJ, Woo SLC, Smith LC;
DR WPI, 2000-281993/24.
XX
XX
XX System for transporting nucleic acid into cells, useful e.g. in gene
PT therapy and for generating transgenic animals, comprises binding agent
PT linked to nucleic acid, surface ligand and lytic agent -
XX
XX
XX Disclosure; Column 125-128; 108pp; English.
XX
XX
XX The present invention relates to a transporter system for delivering
CC nucleic acid to a cell. The system comprises a nucleic acid binding
CC complex, consisting of a binding molecule bonded non-covalently to the
CC nucleic acid, and covalently to a surface ligand, and a lytic agent. The
CC binding molecule is spermine or a spermidine derivative. Nucleotide
CC sequences AAA3633-36652 and peptide sequences AA98456-998500 are used
CC in the construction of the transporter system of the invention. The
CC transporter system is used in gene therapy, particularly to deliver
CC nucleic acids to hepatocytes, muscle cells or bone forming cells, e.g
CC for treating cardiovascular disease, cancer, and infection. The
CC transporter systems are also used to create transgenic animals (as models
CC for human carcinogenesis or disease or for drug testing). Other uses
CC include transforming cells to produce proteins, or transfecting cells in
CC vitro to study the function of the nucleic acid. The use of a surface
CC ligand allows specific targeting of selected cells and tissues. The lytic
CC agent provides for release of the nucleic acid into the cellular
CC interior, from endosomes, without requiring endosomal or lysosomal

```


Job time : 53.3766 sec

QY 57 KK--EAYKAEAKKAYAKKAEKKEA-----AABAKKAEAKKAYK-AEAKKAAAEAA 106
 DB 505 TKVAEAEKQKAEATKVAEAEKQKAEATKVAEAEKQKAEATKVAEAEKQKAAABA 561

RESULT 15

AAV34068
 ID AAV34068 standard; peptide: 158 AA.

XX AAV34068;

XX AC
 XX AAV34068;

XX DT
 XX 23-NOV-1999 (first entry)

XX DE
 XX Histone H1 isoform H1.5 PANCA-reactive fragment (residues 69-226).

XX KW
 KW Ulcerative colitis; histone; H1-like antigen; porin antigen; human;

XX KM
 KM Bacteroides antigen; inflammatory bowel disease; IBD; PANCA; diagnosis;

XX KM
 KM perinuclear anti-neutrophil cytoplasmic antibody; isoform.

XX OS
 OS Homo sapiens.

XX PN
 PN MO9945955-A1.

XX PD
 PD 16-SEP-1999.

XX PF
 PF 12-MAR-1999; 99WO-US05492.

XX PR
 PR 12-MAR-1998; 98US-0041889.

XX PA
 PA (REGC) UNIV CALIFORNIA.

XX PI
 PI Braun J, Cohavy O;

XX DR
 DR WPI; 1999-551215/46.

XX PT
 PT Use of histone H1, porin or Bacteroides antigens as targets for the

XX PT
 PT diagnosis, prevention and treatment of ulcerative colitis -

XX PS
 PS Example 4; Page 125-126; 134pp; English.

XX CC
 CC The invention provides a method for the diagnosis, prevention and

XX CC
 CC treatment of ulcerative colitis (UC) using histone H1-like antigen, a

XX CC
 CC porin antigen or a Bacteroides antigen as a target antigen. The novel

XX CC
 CC method of diagnosing UC in a subject suspected of having inflammatory

XX CC
 CC bowel disease (IBD) comprises: (1) obtaining a sample from the subject;

XX CC
 CC (2) contacting the sample with a histone H1-like antigen, or perinuclear

XX CC
 CC anti-neutrophil cytoplasmic antibody (PANCA)-reactive fragment, to form a

XX CC
 CC complex of the histone H1-like antigen, or the PANCA-reactive fragment,

XX CC
 CC and antibody to the histone H1-like antigen; and (3) detecting the

XX CC
 CC presence or absence of the complex; where the presence of the complex

XX CC
 CC indicates that the subject has UC. The PANCA-reactive histone H1-like

XX CC
 CC antigen, porin antigen and Bacteroides antigen are useful in the

XX CC
 CC diagnosis, prevention and treatment of UC. The methods can also be used

XX CC
 CC for identifying agents useful for treating UC. The present sequence

XX CC
 CC represents a human histone H1 isoform H1.5 PANCA-reactive fragment.

XX SQ
 SQ Sequence 158 AA;

XX Query Match 30.9%; Score 160.5; DB 20; Length 158;

XX Best Local Similarity 46.8%; Pred No. 2.2e-06;

XX Matches 51; Conservative 7; Mismatches 44; Indels 7; Gaps 2;

XX QY 1 AKKYAKKAEAKKAYAKKAEKKEA-----EAKKKAEEAKKAYKAEAAK 55

XX DB 51 AKPKAKKAGAAKAKKAGATPKK--AKKAAGAKKAVKTKPKKAKKPAAGVKVAKSPK 108

XX QY 56 AKKEAYKAEAKKAYKAAAEKKEGYAAAEAKKAEAAKAYKAEAAKAAKE 104

XX DB 109 AKAAKPKKATKSPAKPKAVKPKAKPKAKPKAKPKAKPKAKPKAKKAAK 157

Search completed: January 21, 2004, 08:59:10


```

RESULT 7
US-10-205-979-52
; Sequence 52, Application US/10205979
; Publication No. US20030147861A1
; GENERAL INFORMATION:
; APPLICANT: Watson, James D.
; APPLICANT: Tan, Paul L. J.
; APPLICANT: Abernethy, Nevyn
; TITLE OF INVENTION: Compounds and Methods for the Modulation
; TITLE OF INVENTION: of Immune Responses
; FILE REFERENCE: 11000.1063U
; CURRENT APPLICATION NUMBER: US/10/205,979
; CURRENT FILING DATE: 2002-07-25
; PRIOR APPLICATION NUMBER: 60/308,446
; PRIOR FILING DATE: 2001-07-26
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 52
; LENGTH: 223
; TYPE: prt
; ORGANISM: Mycobacterium vaccae
US-10-205-979-52

```

```

RESULT 9
US/10-156-761-9889
, Sequence 9889, Application US/10156761
, Publication No. US20030119018A1
, GENERAL INFORMATION:
, APPLICANT: OMURA, SATOSHI
, APPLICANT: IKEDA, HARUO
, APPLICANT: ISHIKAWA, JUN
, APPLICANT: HORIKAWA, HIROSHI
, APPLICANT: SHIBA, TADAYOSHI
, APPLICANT: SAKAKI, YOSHIYUKI
, APPLICANT: HATTORI, MASAIIRA
, TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
, FILE REFERENCE: 249-262
, CURRENT APPLICATION NUMBER: US/10/156,761
, CURRENT FILING DATE: 2002-05-29
, PRIOR APPLICATION NUMBER: JP 2001-204089
, PRIOR FILING DATE: 2001-05-30
, PRIOR APPLICATION NUMBER: JP 2001-272697
, PRIOR FILING DATE: 2001-08-02
, NUMBER OF SEQ ID NOS: 15109
, SEQ ID NO 9889

```

LENGTH: 376
TYPE: PRT
ORGANISM: Streptomyces avermitilis
US-10-156-761-9889

Query Match 31.4%; Score 163; DB 15; Length 376;
Best Local Similarity 44.9%; Pred. No. 9.9e-06;
Matches 48; Conservative 15; Mismatches 38; Indels 6; Gaps 2;

QY 6 KKAAYAKKAAKAAK---KKAAYAKKAAKAAK---KKAAYAKKAAKAAK 59
DB 245 RAKKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAK 304

QY 60 AKKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAK 106
DB 305 AKKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAK 351

RESULT 10
US-09-820-843A-27
Sequence 27, Application US/09820843A
Publication No. US20030039963A1
GENERAL INFORMATION:

APPLICANT: Council of Scientific and Industrial Research
TITLE OF INVENTION: A COMPUTATIONAL METHOD FOR THE IDENTIFICATION OF CANDIDATE PROTEI
FILE REFERENCE: Q63915
CURRENT APPLICATION NUMBER: US/09/820,843A
CURRENT FILING DATE: 2001-03-30
NUMBER OF SEQ ID NOS: 118
SOFTWARE: Patent version 3.0
SEQ ID NO 27
LENGTH: 356
TYPE: PRT
ORGANISM: Vibrio cholerae
FEATURE:
NAME/KEY: misc.feature
OTHER INFORMATION: coli protein
NAME/KEY: misc.feature
OTHER INFORMATION: gii9656364
US-09-820-843A-27

Query Match 31.2%; Score 162; DB 11; Length 356;
Best Local Similarity 42.6%; Pred. No. 1.1e-05;
Matches 55; Conservative 13; Mismatches 37; Indels 24; Gaps 5;

QY 3 KKAAYAKKAAKAAK---AKAAYAKKAAKAAKAAKAAK---AKAAYAKKAAKAAK 56
DB 112 KKAAYAKKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAK 171
QY 57 KKAAYAKKAAKAAK---AKAAYAKKAAKAAKAAKAAKAAKAAKAAKAAK 100
DB 172 KKAAYAKKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAK 229
QY 101 AKKAAAYEA 109
DB 230 EAERKAAEA 238

RESULT 11
US-10-229-567-40
Sequence 40, Application US/10229567
Publication No. US20030092080A1
GENERAL INFORMATION:

APPLICANT: Braun, Jonathan
TITLE OF INVENTION: Diagnosis, Prevention and Treatment of
Ulcerative Colitis, and Clinical Subtypes Thereof, Using
Microbial UC PANCA antigens
NUMBER OF SEQUENCES: 41
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Campbell & Flores LLP
STREET: 4370 La Jolla Village Drive, Suite 700

CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/229,567
FILING DATE: 27-Aug-2002
CLASSIFICATION: <unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/417,264
FILING DATE: <unknown>
APPLICATION NUMBER: US 09/041,889
FILING DATE: <unknown>

ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-PW 3006
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949

INFORMATION FOR SEQ ID NO: 40:
SEQUENCE CHARACTERISTICS:
LENGTH: 158 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 40:
US-10-229-567-40

Query Match 30.9%; Score 160.5; DB 15; Length 158;
Best Local Similarity 46.8%; Pred. No. 6.4e-06;
Matches 51; Conservative 7; Mismatches 44; Indels 7; Gaps 2;

QY 1 AKKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAK 55
DB 51 AKKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAK 108

QY 56 AKKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAK 104
DB 109 AKKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAKAAK 157

RESULT 12
US-10-229-567-32
Sequence 32, Application US/10229567
Publication No. US20030092080A1
GENERAL INFORMATION:

APPLICANT: Braun, Jonathan
TITLE OF INVENTION: Diagnosis, Prevention and Treatment of
Ulcerative Colitis, and Clinical Subtypes Thereof, Using
Microbial UC PANCA antigens
NUMBER OF SEQUENCES: 41
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Campbell & Flores LLP
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/229,567
FILING DATE: 27-Aug-2002

CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/417,264
FILING DATE: <Unknown>
APPLICATION NUMBER: US 09/041,889
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-PM 3006
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 32:
SEQUENCE CHARACTERISTICS:
LENGTH: 226 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 32:
US-10-229-567-32

Query Match 30.9%; Score 160.5; DB 15; Length 226;
Best Local Similarity 46.8%; Pred. No. 9.3e-06;
Matches 51; Conservative 7; Mismatches 44; Indels 7; Gaps 2;

QY 1 AKYAKAEKAYAKAKAEKKAYAKAEKAYAKA-----BAKKAYAEAKTKAEAK 55
DB 119 AKPKAKGAKAKAKKPAKATPKK--AKKAGAKKAVKKTTPKAKKPAAGVKVAKSPK 176

QY 56 AKKAYKAEKAYAKAKAEKKAYAKAEKAYAKAEKAYAKAEKAYAKAEKAYAKAE 104
DB 177 AKAAKPKKATKSPKPKAVPKAKPKPKAKPKAKPKAKPKAKPKAKPKAKPKAKK 225

RESULT 13
US-10-262-209-1
Sequence 1, Application US/10262209
Publication No. US20030125239A1
GENERAL INFORMATION:
APPLICANT: Criaenti, Andrea
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-PM 3006
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 32:
SEQUENCE CHARACTERISTICS:
LENGTH: 226 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 32:
US-10-262-209-1

Query Match 30.7%; Score 159.5; DB 15; Length 234;
Best Local Similarity 45.7%; Pred. No. 1.2e-05;
Matches 53; Conservative 8; Mismatches 44; Indels 11; Gaps 3;

QY 2 KKYAKAEKAYAKAKAEKK--AYAKAEKAYAKAEKKAYAKAEKAYAKAEKAYAKAE 59
DB 109 KKAAGGKPKKAKKAGAKKPAKATPKKATPKKATPKKATPKKATPKKATPKKATPK 166

QY 60 AYKAEKAYAKAKAEKAYAKAEKAYAKAEKAYAKAEKAYAKAEKAYAKAEKAYAK 108
DB 167 AKKAKSPKAKAKKPKAKPKSPAKAKAVPKAKAKTKAKPKAKKPKAKKPKAKK 222

RESULT 14
US-10-229-567-4
Sequence 4, Application US/10229567
Publication No. US20030092080A1
GENERAL INFORMATION:
APPLICANT: Braun, Jonathan
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-PM 3006
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 218 amino acids
TYPE: amino acid
TOPOLOGY: linear
FEATURE:
NAME/KEY: Peptide
LOCATION: 1..218
OTHER INFORMATION: /note= "product = Human Histone
H1-8-4"
SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-10-229-567-4

Query Match 30.4%; Score 158; DB 15; Length 218;
Best Local Similarity 48.6%; Pred. No. 1.4e-05;
Matches 53; Conservative 7; Mismatches 43; Indels 6; Gaps 3;

QY 2 KKYAKAEKAYAKAKAEKK--AYAKAEKAYAKAEKKAYAKAEKAYAKAEKAYAKAE 59
DB 108 KKAAGGKPKKAKKAGAKKPAKATPKKATPKKATPKKATPKKATPKKATPKKATPK 165

QY 60 AYKAEKAYAKAEKAYAKAEKAYAKAEKAYAKAEKAYAKAEKAYAKAEKAYAKAE 106
DB 166 AKKAKSPKAKAKKPKAKPKSPAKAKAVPKAKAKTKAKPKAKKPKAKKPKAKK 214

RESULT 15
US-10-262-209-2
Sequence 2, Application US/10262209
Publication No. US20030125239A1
GENERAL INFORMATION:
APPLICANT: Criaenti, Andrea

APPLICANT: Essegbir, Selma
TITLE OF INVENTION: Compositions for Drug Delivery
FILE REFERENCE: GJB-6703
CURRENT APPLICATION NUMBER: US/10/262,209
CURRENT FILING DATE: 2002-09-30
PRIOR APPLICATION NUMBER: UK 0218324.2
PRIOR FILING DATE: 2002-08-07
PRIOR APPLICATION NUMBER: PCT/GB01/01699
PRIOR FILING DATE: 2001-04-12
PRIOR APPLICATION NUMBER: UK 0102667.3
PRIOR FILING DATE: 2001-02-02
PRIOR APPLICATION NUMBER: UK 0009080.3
PRIOR FILING DATE: 2000-04-12
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn version 3.1
SEQ ID NO 2
LENGTH: 130
TYPE: PRT
ORGANISM: Homo sapiens
US-10-262-209-2

Query Match 29.8%; Score 154.5; DB 15; Length 130;
Best Local Similarity 47.3%; Pred. No. 1.6e-05;
Matches 52; Conservative 8; Mismatches 43; Indels 7; Gaps 3;

QY 2 KKYAKKAEKAVAKK--AAAKEKAYAKKEAKAYKAAEAKKAKAEAKKYAKKEAAKAKK 58
DB 17 KKAASGEAKPKAKKGAAPAAKPPKATGATPPKSAKTPKAKKPPAAAA-- 74
QY 59 EAYKAEAKKYAKAEK--KEVAAAEAKKAEAKAYKAEAAKAAKAA 106
DB 75 GKAKKSPKAAKPKAKPSPAKAKAVKPAKPKTKPAKPKKAA 124

Search completed: January 21, 2004, 09:10:08
Job time : 39.6364 secs


```
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
OTHER INFORMATION: PEPTIDE
US-09-405-743A-6

Query Match      60.2%; Score 312.5; DB 4; Length 86;
Best Local Similarity 72.1%; Pred. No. 6.5e-20;
Matches 80; Conservative 1; Mismatches 3; Indels 27; Gaps 5;

QY 1 AKKYAKKAEKAYAKKAEKAYAKKAEKAYAKKAEKAYAKKAEKAYAKKAEKAE 60
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1 AKKYAKK-EKAYAKKA-----EKAKKAEKAYAKKAEKAEKAEKAEKAEKAEKAE 36

QY 61 YKAEAKKYAKAKAEKKEKYAAAEAK-KAEAA-KAYAKAEAKAKAEAYEA 109
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 37 -KAEAKKYAKAKAEKKEKYAAAEAK-----YKAEAAKAEKAEAYEA 86

RESULT 3
US-09-405-743A-5
Sequence 5, Application US/09405743A
Patent No. 6514938
GENERAL INFORMATION:
APPLICANT: Yeda Research and Development Co., Ltd.
TITLE OF INVENTION: GLATIRAMER ACETATE MOLECULAR WEIGHT MARKERS
FILE REFERENCE: 60807-A
CURRENT APPLICATION NUMBER: US/09/405,743A
CURRENT FILING DATE: 1999-09-24
NUMBER OF SEQ ID NOS: 7
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 5
LENGTH: 77
TYPE: PRF
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
OTHER INFORMATION: PEPTIDE
US-09-405-743A-5

Query Match      55.7%; Score 289; DB 4; Length 77;
Best Local Similarity 67.0%; Pred. No. 5.4e-18;
Matches 73; Conservative 1; Mismatches 3; Indels 32; Gaps 4;

QY 1 AKKYAKKAEKAYAKKAEKKEKYAAAEAKKAEAKKAYKAEAAKAEAYEA 109
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1 AKKYAKK-EKAYAKKA-----EKAKKAEKAEKAYKAEAAKAEAYEA 36

QY 61 YKAEAKKYAKAKAEKKEKYAAAEAKKAEAKKAYKAEAAKAEAYEA 109
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 37 -KAEAKKYAKAKAEKKEKYAAAEAK-----YKAEAAKAEKAEAYEA 77

RESULT 4
US-09-405-743A-4
Sequence 4, Application US/09405743A
Patent No. 6514938
GENERAL INFORMATION:
APPLICANT: Yeda Research and Development Co., Ltd.
TITLE OF INVENTION: GLATIRAMER ACETATE MOLECULAR WEIGHT MARKERS
FILE REFERENCE: 60807-A
CURRENT APPLICATION NUMBER: US/09/405,743A
CURRENT FILING DATE: 1999-09-24
NUMBER OF SEQ ID NOS: 7
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 4
LENGTH: 66
TYPE: PRF
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
OTHER INFORMATION: PEPTIDE
US-09-405-743A-4
```

```
Query Match      44.0%; Score 228.5; DB 4; Length 66;
Best Local Similarity 56.9%; Pred. No. 5.2e-13;
Matches 62; Conservative 0; Mismatches 4; Indels 43; Gaps 4;

QY 1 AKKYAKKAEKAYAKKAEKKEKYAAAEAKKAEAKKAYKAEAAKAEAYEA 60
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1 AKKYAKK-EKAYAKKA-----EKAKKAEKAEKAYKAEAAKAEAYEA 25

QY 61 YKAEAKKYAKAKAEKKEKYAAAEAKKAEAKKAYKAEAAKAEAYEA 109
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 26 -KAEAKKYAKAKAEKKEKYAAAEAK-----YKAEAAKAEKAEAYEA 66

RESULT 5
US-09-405-743A-3
Sequence 3, Application US/09405743A
Patent No. 6514938
GENERAL INFORMATION:
APPLICANT: Yeda Research and Development Co., Ltd.
TITLE OF INVENTION: GLATIRAMER ACETATE MOLECULAR WEIGHT MARKERS
FILE REFERENCE: 60807-A
CURRENT APPLICATION NUMBER: US/09/405,743A
CURRENT FILING DATE: 1999-09-24
NUMBER OF SEQ ID NOS: 7
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 3
LENGTH: 56
TYPE: PRF
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC
OTHER INFORMATION: PEPTIDE
US-09-405-743A-3

Query Match      34.8%; Score 180.5; DB 4; Length 56;
Best Local Similarity 45.9%; Pred. No. 4.5e-09;
Matches 50; Conservative 3; Mismatches 3; Indels 53; Gaps 3;

QY 1 AKKYAKKAEKAYAKKAEKKEKYAAAEAKKAEAKKAYKAEAAKAEAYEA 60
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1 AKKYAKK-EKAYAKKA-----EKAKKAEKAEKAYKAEAAKAEAYEA 39

QY 61 YKAEAKKYAKAKAEKKEKYAAAEAKKAEAKKAYKAEAAKAEAYEA 109
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 40 -----YKAEAAKAEKAEAYEA 56

RESULT 6
US-08-460-890A-64
Sequence 64, Application US/08460890A
Patent No. 5994109
GENERAL INFORMATION:
APPLICANT: MCO, Savio L.C.
APPLICANT: Smith, Louis C.
APPLICANT: Cristiano, Richard J.
APPLICANT: Gotchaik, Stephen
TITLE OF INVENTION: NUCLEIC ACID TRANSPORTER SYSTEMS AND
NUMBER OF SEQUENCES: 65
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon
STREET: 633 West Fifth Street
CITY: Los Angeles
STATE: California
COUNTRY: U.S.A.
ZIP: 90071-2066
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 MB
MEDIUM TYPE: storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: IBM P.C. DOS 5.0
SOFTWARE: FastSeq for Windows 2.0
```

```

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/855,389
FILING DATE: March 20, 1992
APPLICATION NUMBER: PCT/US93/02725
FILING DATE: March 19, 1993
ATTORNEY/AGENT INFORMATION:
NAME: Warburg, Richard J.
REGISTRATION NUMBER: 32,327
REFERENCE/DOCKET NUMBER: 205/012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (213) 489-1600
TELEFAX: (213) 955-0440
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 64:
SEQUENCE CHARACTERISTICS:
LENGTH: 100 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
OTHER INFORMATION: "Lys Ala" in positions 3 to 100 may be
OTHER INFORMATION: present or absent.
US-08-167-641C-64

Query Match      34.6%; Score 179.5; DB 3; Length 100;
Best Local Similarity 55.0%; Pred. No. 9.7e-09;
Matches 55; Conservative 8; Mismatches 34; Indels 3; Gaps 3

Cy      6 KKAERAKAK-KAKAAKKKAYAKKEAVYKAAAEAKKKAKAEKKYAKAEAAVAKKEAYNAE 64
        ||| ||||| : ||| : ||| : ||| ||| : ||| : ||| : ||| : ||| :
Dd      1 KAKKAAKAKAKAKKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKA-KAK 59
        ||| ||||| : ||| : ||| : ||| ||| : ||| : ||| : ||| : ||| :
Oy      65 AKKYAKA-AKAKEVEYAAAEAKKAEAAKAYKAEAKAAK 103
        ||| ||||| : ||| : ||| : ||| ||| : ||| : ||| : ||| : ||| :
Db      60 AKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAKAK 99
        ||| ||||| : ||| : ||| : ||| ||| : ||| : ||| : ||| : ||| :

RESULT 8
US-08-460-971A-64
Sequence 64, Application US/08460971A
Patent No. 6150168
GENERAL INFORMATION:
APPLICANT: Woo, Savio L.C.
APPLICANT: Smith, Louis C.
APPLICANT: Cristiano, Richard J.
APPLICANT: Gottchalk, Stephen
TITLE OF INVENTION: NUCLEIC ACID TRANSPORTER SYSTEMS AND
NUMBER OF SEQUENCES: 65
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon
STREET: 633 West Fifth Street
CITY: Suite 4700
STATE: Los Angeles
COUNTRY: California
ZIP: 90071-2066
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
MEDIUM TYPE: Storage
COMPUTER: IBM Compatible
OPERATING SYSTEM: IBM P.C. DOS 5.0
SOFTWARE: FastSeq for Windows 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/460,971A
FILING DATE: June 5, 1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/167,641
FILING DATE: December 14, 1993
APPLICATION NUMBER: 07/855,389

```

[illegible]

APPLICANT: Tan, Paul
APPLICANT: Visser, Elizabeth
APPLICANT: Skinner, Margot
APPLICANT: Preestidge, Rose
TITLE OF INVENTION: Compounds and Methods for
TITLE OF INVENTION: Treatment and Diagnosis of Mycobacterial Infections
NUMBER OF SEQUENCES: 208
CORRESPONDENCE ADDRESS:
ADDRESSEE: Law Offices of Ann W. Speckman
STREET: 2601 Elliott Avenue, Suite 4185
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98121
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: PaeSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/095,855
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/705,347
FILING DATE: 29-AUG-1996
APPLICATION NUMBER: 08/873,970
FILING DATE: 12-JUN-1997
APPLICATION NUMBER: 08/997,362
FILING DATE: 23-DEC-1997
ATTORNEY/AGENT INFORMATION:
NAME: Sleath, Janet
REGISTRATION NUMBER: 37,007
REFERENCE/DOCKET NUMBER: 11000.1002C3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-269-0565
TELEFAX: 206-269-0563
TELEX:
INFORMATION FOR SEQ ID NO: 201:
SEQUENCE CHARACTERISTICS:
LENGTH: 223 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-095-855-201
Query Match 34.3%; Score 178; DB 3; Length 223;
Best Local Similarity 54.5%; Pred. No. 2.9e-08;
Matches 60; Conservative 6; Mismatches 34; Indels 10; Gaps 5;
Qy 1 AKKYAKK--AEKAYAKKAKAKE--KKAYAKKEKAYKAAEAKKKAKAEKAYKAEAK 55
Db 112 ARKAKKAPAKKAAKAAKAAKAPAKKA-ATKAPAKKATPAKKAAGAAKATAKKAP 170
Qy 56 AKKEAYKAEKKYKAAKAEKKEYA--AAEKKAAKAYKAEKAAK 103
Db 171 AKKAPAKKATKAAAPAKKAPAKKATKAAAPAKKAPAK--KAPAKKAPAK 218
RESULT 12
US-09-205-426-201
Sequence 201, Application US/09205426
Patent No. 6406704
GENERAL INFORMATION:
APPLICANT: Watson, James D.
APPLICANT: Tan, Paul L. J.
TITLE OF INVENTION: Compounds and Methods for Treatment and
TITLE OF INVENTION: Diagnosis of Mycobacterial Infections
FILE REFERENCE: 11000.1002C4
CURRENT APPLICATION NUMBER: US/09/205,426
CURRENT FILING DATE: 1998-12-04
EARLIER APPLICATION NUMBER: 09/095,855

EARLIER FILING DATE: 1998-06-11
EARLIER APPLICATION NUMBER: 08/997,362
EARLIER FILING DATE: 1997-12-23
EARLIER APPLICATION NUMBER: 08/873,970
EARLIER FILING DATE: 1997-06-12
EARLIER APPLICATION NUMBER: 08/705,347
EARLIER FILING DATE: 1996-08-29
NUMBER OF SEQ ID NOS: 208
SOFTWARE: PaeSeq for Windows Version 3.0
SEQ ID NO 201
LENGTH: 223
TYPE: PRT
ORGANISM: Mycobacterium vaccae
US-09-205-426-201
Query Match 34.3%; Score 178; DB 4; Length 223;
Best Local Similarity 54.5%; Pred. No. 2.9e-08;
Matches 60; Conservative 6; Mismatches 34; Indels 10; Gaps 5;
Qy 1 AKKYAKK--AEKAYAKKAKAKE--KKAYAKKEKAYKAAEAKKKAEKAYKAEAK 55
Db 112 ARKAKKAPAKKAAKAAKAAKAPAKKA-ATKAPAKKATPAKKAAPAKKATPAKKAAP 170
Qy 56 AKKEAYKAEKKYKAAKAEKKEYA--AAEKKAAKAYKAEKAAK 103
Db 171 AKKAPAKKATKAAAPAKKAPAKKATKAAAPAKKAPAK--KAPAKKAPAK 218
RESULT 13
US-08-216-894-8
Sequence 8, Application US/08216894
Patent No. 5876734
GENERAL INFORMATION:
APPLICANT: Kirchhoff, Louis V.
APPLICANT: Otsu, Keiko
TITLE OF INVENTION: POLYPEPTIDES FOR DIAGNOSING INFECTION
TITLE OF INVENTION: WITH TRYPAKOSOMA CRUZI
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington, D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/216,894
FILING DATE: 24-MAR-1994
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 85326/102/DRL0
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 672-5300
TELEFAX: (202) 672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 643 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-216-894-8
Query Match 31.3%; Score 162.5; DB 2; Length 643;
Best Local Similarity 44.4%; Pred. No. 1.6e-06;
Matches 52; Conservative 18; Mismatches 36; Indels 11; Gaps 4;
Qy 1 AKKYAKKAEKKYKAAKAEKKEKAYKAEKAYKAEKKEKAAKAAKAAKAAKAAKAAK 56

